(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 April 2003 (10.04.2003)

PCT

(10) International Publication Number
WO 03/029423 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/31358

(22) International Filing Date: 2 October 2002 (02.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/326,483	2 October 2001 (02.10.2001)	US
60/327,342	5 October 2001 (05.10.2001)	US
60/328,029	9 October 2001 (09.10.2001)	US
60/328,056	9 October 2001 (09.10.2001)	US
60/328,044	9 October 2001 (09.10.2001)	US
60/327,917	9 October 2001 (09.10.2001)	US
60/328,849	12 October 2001 (12.10.2001)	US
60/329,414	15 October 2001 (15.10.2001)	US
60/330,142	17 October 2001 (17.10.2001)	US
60/341,058	22 October 2001 (22.10.2001)	US
60/343,629	24 October 2001 (24.10.2001)	US
60/339,266	24 October 2001 (24.10.2001)	US
60/349,575	29 October 2001 (29.10.2001)	US
60/346,357	1 November 2001 (01.11.2001)	US
60/371,972	12 April 2002 (12.04.2002)	US
60/371,980	12 April 2002 (12.04.2002)	US
60/373,261	17 April 2002 (17.04.2002)	US
60/373,805	19 April 2002 (19.04.2002)	US
60/374,738	23 April 2002 (23.04.2002)	US
60/381,101	16 May 2002 (16.05.2002)	US
60/381,635	17 May 2002 (17.05.2002)	US
60/383,830	29 May 2002 (29.05.2002)	US
10/262,839	1 October 2002 (01.10.2002)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/326,483 (CIP)
Filed on	2 October 2001 (02.10.2001)
US	60/327,342 (CIP)
Filed on	5 October 2001 (05.10.2001)
US	60/327,917 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,029 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,056 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,044 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,849 (CIP)
Filed on	12 October 2001 (12.10.2001)
US	60,329,414 (CIP)
Filed on	15 October 2001 (15.10.2001)
US	60/330,142 (CIP)
Filed on	17 October 2001 (17.10.2001)
US	60/341,058 (CIP)
Filed on	22 October 2001 (22.10.2001)
US	60/343,629 (CIP)
Filed on	24 October 2001 (24.10.2001)
US	60/339,266 (CIP)
Filed on	24 October 2001 (24.10.2001)
US	60/349,575 (CIP)
Filed on	29 October 2001 (29.10.2001)
US	60/346,357 (CIP)
Filed on	1 November 2001 (01.11.2001)
US	60/371,972 (CIP)
Filed on	12 April 2002 (12.04.2002)
US	60/371,980 (CIP)
Filed on	12 April 2002 (12.04.2002)
US	60/373,261 (CIP)
Filed on	17 April 2002 (17.04.2002)
US	60/373,805 (CIP)
Filed on	19 April 2002 (19.04.2002)
US	60/374,738 (CIP)
Filed on	23 April 2002 (23.04.2002)
US	60/381,101 (CIP)
Filed on	16 May 2002 (16.05.2002)
US	60/381,635 (CIP)
Filed on	17 May 2002 (17.05.2002)
US	60/383,830 (CIP)
Filed on	29 May 2002 (29.05.2002)
US	10/262,839 (CIP)
Filed on	1 October 2002 (01.10.2002)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ALSOBROOK, John, P., II** [US/US]; 60 Lake Drive, Madison, CT 06443

[Continued on next page]

(54) Title: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies that immunospecifically bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the novel polypeptide, polynucleotide, or antibody specific to the polypeptide. Vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using same are also included. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

WO 03/029423 A2

(US). **ANDERSON, David, W.** [US/US]; 85 Montoya Drive, Branford, CT 06405 (US). **BOLDOG, Ferenc, L.** [HU/US]; 1687 Hartford Turnpike, North Haven, CT 06473 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). **CATTERTON, Elina** [FI/US]; 584 Boston Post Road, Madison, CT 06443 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **ELLERMAN, Karen** [US/US]; 87 Montoya Drive, Branford, CT 06405 (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **JI, Weizhen** [CN/US]; 375 Old Rock Road, Branford, CT 06405 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US). **LEACH, Martin, D.** [GB/US]; 131 Beechwood Drive, Madison, CT 06433 (US). **LI, Li** [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **RIEGER, Daniel, K.** [DE/US]; 10A McKinnel Court, Branford, CT 06405 (US). **ROTHENBERG, Mark, E.** [US/US]; 2 Allen Road, Clinton, CT 06413 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Indian Meadows Drive, Guilford, Connecticut 06437 (US). **SMITHSON, Glennda** [US/US]; 125 Michael Drive, Guilford, CT 06435 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, Number 1, New Haven, CT 06511 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **VERNET, Corine, A., M.** [FR/US]; 1739 Foxon Road, Apartment L6, Branford, CT 06471 (US). **VOSS, Edward, Z.** [US/US]; 123 Knollwood Drive, Wallingford, CT 06492 (US). **ZERHUSEN, Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **ZHONG, Mei** [CA/US]; 45 Harrison Avenue, Apartment 1B, Branford, CT 06405 (US).

(74) Agents: ELRIFI, Ivor, R. et al.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS OF USE

FIELD OF THE INVENTION

The present invention relates to novel polypeptides, and the nucleic acids encoding them, having properties related to stimulation of biochemical or physiological responses in a cell, a tissue, an organ or an organism. More particularly, the novel polypeptides are gene products of novel genes, or are specified biologically active fragments or derivatives thereof. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

BACKGROUND OF THE INVENTION

Eukaryotic cells are characterized by biochemical and physiological processes which under normal conditions are exquisitely balanced to achieve the preservation and propagation of the cells. When such cells are components of multicellular organisms such as vertebrates, or more particularly organisms such as mammals, the regulation of the biochemical and physiological processes involves intricate signaling pathways. Frequently, such signaling pathways involve extracellular signaling proteins, cellular receptors that bind the signaling proteins, and signal transducing components located within the cells.

Signaling proteins may be classified as endocrine effectors, paracrine effectors or autocrine effectors. Endocrine effectors are signaling molecules secreted by a given organ into the circulatory system, which are then transported to a distant target organ or tissue. The target cells include the receptors for the endocrine effector, and when the endocrine effector binds, a signaling cascade is induced. Paracrine effectors involve secreting cells and receptor cells in close proximity to each other, for example two different classes of cells in the same tissue or organ. One class of cells secretes the paracrine effector, which then reaches the second class of cells, for example by diffusion through the extracellular fluid. The second class of cells contains the receptors for the paracrine effector; binding of the effector results in induction of the signaling cascade that elicits the corresponding biochemical or physiological effect. Autocrine effectors are highly analogous to paracrine effectors, except that the same cell type that secretes the autocrine effector also contains the receptor. Thus the autocrine effector binds to receptors on the same cell, or on identical neighboring cells. The binding process then elicits the characteristic biochemical or physiological effect.

Signaling processes may elicit a variety of effects on cells and tissues including by way of nonlimiting example induction of cell or tissue proliferation, suppression of growth or proliferation, induction of differentiation or maturation of a cell or tissue, and suppression of differentiation or maturation of a cell or tissue.

Many pathological conditions involve dysregulation of expression of important effector proteins. In certain classes of pathologies the dysregulation is manifested as diminished or suppressed level of synthesis and secretion of protein effectors. In other classes of pathologies the dysregulation is manifested as increased or up-regulated level of synthesis and secretion of protein effectors. In a clinical setting a subject may be suspected

of suffering from a condition brought on by altered or mis-regulated levels of a protein effector of interest. Therefore there is a need to assay for the level of the protein effector of interest in a biological sample from such a subject, and to compare the level with that characteristic of a nonpathological condition. There also is a need to provide the protein effector as a product of manufacture. Administration of the effector to a subject in need thereof is useful in treatment of the pathological condition. Accordingly, there is a need for a method of treatment of a pathological condition brought on by a diminished or suppressed levels of the protein effector of interest. In addition, there is a need for a method of treatment of a pathological condition brought on by a increased or up-regulated levels of the protein effector of interest.

Antibodies are multichain proteins that bind specifically to a given antigen, and bind poorly, or not at all, to substances deemed not to be cognate antigens. Antibodies are comprised of two short chains termed light chains and two long chains termed heavy chains. These chains are constituted of immunoglobulin domains, of which generally there are two classes: one variable domain per chain, one constant domain in light chains, and three or more constant domains in heavy chains. The antigen-specific portion of the immunoglobulin molecules resides in the variable domains; the variable domains of one light chain and one heavy chain associate with each other to generate the antigen-binding moiety. Antibodies that bind immunospecifically to a cognate or target antigen bind with high affinities. Accordingly, they are useful in assaying specifically for the presence of the antigen in a sample. In addition, they have the potential of inactivating the activity of the antigen.

Therefore there is a need to assay for the level of a protein effector of interest in a biological sample from such a subject, and to compare this level with that characteristic of a nonpathological condition. In particular, there is a need for such an assay based on the use of an antibody that binds immunospecifically to the antigen. There further is a need to inhibit the activity of the protein effector in cases where a pathological condition arises from elevated or excessive levels of the effector based on the use of an antibody that binds immunospecifically to the effector. Thus, there is a need for the antibody as a product of manufacture. There further is a need for a method of treatment of a pathological condition brought on by an elevated or excessive level of the protein effector of interest based on administering the antibody to the subject.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of isolated polypeptides including amino acid sequences selected from mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1, NOV2, NOV3, *etc.*, nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

The invention also is based in part upon variants of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed. In another embodiment, the invention includes the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107. In another embodiment, the invention also comprises variants of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed. The invention also involves fragments of any of the mature forms of the amino acid sequences selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107, or any other amino acid sequence selected from this group. The invention also comprises fragments from these groups in which up to 15% of the residues are changed.

In another embodiment, the invention encompasses polypeptides that are naturally occurring allelic variants of the sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107. These allelic variants include amino acid sequences that are the translations of nucleic acid sequences differing by a single nucleotide from nucleic acid sequences selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 107. The variant polypeptide where any amino acid changed in the chosen sequence is changed to provide a conservative substitution.

In another embodiment, the invention comprises a pharmaceutical composition involving a polypeptide with an amino acid sequence selected from the group consisting of

SEQ ID NO:2n, wherein n is an integer between 1 and 107 and a pharmaceutically acceptable carrier. In another embodiment, the invention involves a kit, including, in one or more containers, this pharmaceutical composition.

In another embodiment, the invention includes the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease being selected from a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 wherein said therapeutic is the polypeptide selected from this group.

In another embodiment, the invention comprises a method for determining the presence or amount of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 in a sample, the method involving providing the sample; introducing the sample to an antibody that binds immunospecifically to the polypeptide; and determining the presence or amount of antibody bound to the polypeptide, thereby determining the presence or amount of polypeptide in the sample.

In another embodiment, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 in a first mammalian subject, the method involving measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and comparing the amount of the polypeptide in this sample to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, the disease, wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

In another embodiment, the invention involves a method of identifying an agent that binds to a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107, the method including introducing the polypeptide to the agent; and determining whether the agent binds to the polypeptide. The agent could be a cellular receptor or a downstream effector.

In another embodiment, the invention involves a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a polypeptide with an amino

acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107, the method including providing a cell expressing the polypeptide of the invention and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

In another embodiment, the invention involves a method for screening for a modulator of activity or of latency or predisposition to a pathology associated with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107, the method including administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of the invention, wherein the test animal recombinantly expresses the polypeptide of the invention; measuring the activity of the polypeptide in the test animal after administering the test compound; and comparing the activity of the protein in the test animal with the activity of the polypeptide in a control animal not administered the polypeptide, wherein a change in the activity of the polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the polypeptide of the invention. The recombinant test animal could express a test protein transgene or express the transgene under the control of a promoter at an increased level relative to a wild-type test animal The promoter may or may not b the native gene promoter of the transgene.

In another embodiment, the invention involves a method for modulating the activity of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107, the method including introducing a cell sample expressing the polypeptide with a compound that binds to the polypeptide in an amount sufficient to modulate the activity of the polypeptide.

In another embodiment, the invention involves a method of treating or preventing a pathology associated with a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107, the method including administering the polypeptide to a subject in which such treatment or prevention

is desired in an amount sufficient to treat or prevent the pathology in the subject. The subject could be human.

In another embodiment, the invention involves a method of treating a pathological state in a mammal, the method including administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 or a biologically active fragment thereof.

In another embodiment, the invention involves an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107; a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107; a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 or any variant of the polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and the complement of any of the nucleic acid molecules.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

In another embodiment, the invention involves an isolated nucleic acid molecule including a nucleic acid sequence encoding a polypeptide having an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

In another embodiment, the invention comprises an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2n-1, wherein n is an integer between 1 and 107.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 107; a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 107 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 107; and a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 107 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107, wherein the nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group

consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 107, or a complement of the nucleotide sequence.

In another embodiment, the invention includes an isolated nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107, wherein the nucleic acid molecule has a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides in the chosen coding sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

In another embodiment, the invention includes a vector involving the nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107. This vector can have a promoter operably linked to the nucleic acid molecule. This vector can be located within a cell.

In another embodiment, the invention involves a method for determining the presence or amount of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107 in a sample, the method including providing the sample; introducing the sample to a probe that binds to the nucleic acid molecule; and determining the presence or amount of the probe bound to the nucleic acid molecule, thereby determining the presence or amount of the nucleic acid molecule in the sample. The presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type. The cell type can be cancerous.

In another embodiment, the invention involves a method for determining the presence of or predisposition for a disease associated with altered levels of a nucleic acid molecule having a nucleic acid sequence encoding a polypeptide including an amino acid sequence selected from the group consisting of a mature form of the amino acid sequence given SEQ ID NO:2n, wherein n is an integer between 1 and 107 in a first mammalian

subject, the method including measuring the amount of the nucleic acid in a sample from the first mammalian subject; and comparing the amount of the nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease; wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

The invention further provides an antibody that binds immunospecifically to a NOVX polypeptide. The NOVX antibody may be monoclonal, humanized, or a fully human antibody. Preferably, the antibody has a dissociation constant for the binding of the NOVX polypeptide to the antibody less than 1×10^{-9} M. More preferably, the NOVX antibody neutralizes the activity of the NOVX polypeptide.

In a further aspect, the invention provides for the use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, associated with a NOVX polypeptide. Preferably the therapeutic is a NOVX antibody.

In yet a further aspect, the invention provides a method of treating or preventing a NOVX-associated disorder, a method of treating a pathological state in a mammal, and a method of treating or preventing a pathology associated with a polypeptide by administering a NOVX antibody to a subject in an amount sufficient to treat or prevent the disorder.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (amino acid)	Homology
1a	CG105472-01	1	2	Novel KIAA0575/Greb1-like Proteins and Nucleic Acids Encoding Same
2a	CG106287-01	3	4	Integrin Alpha-11
2b	CG106287-02	5	6	Integrin Alpha-11
3a	CG106417-01	7	8	Fibrillin
3b	CG106417-03	9	10	Novel von Willebrand factor
3c	CG106417-04	11	12	Novel von Willebrand factor
3d	209749357	13	14	Fibrillin
3e	CG106417-02	15	16	Fibrillin
4a	CG108901-01	17	18	Cytokine Receptor
4b	CG108901-04	19	20	Cytokine Receptor
4c	CG108901-03	21	22	Cytokine Receptor
4d	CG108901-02	23	24	Cytokine Receptor
5a	CG112505-01	25	26	Laminin 5-Beta 3
5b	CG112505-02	27	28	Laminin 5-Beta 3
6a	CG121965-01	29	30	Fibulin 3
6b	CG121965-02	31	32	Fibulin 3
7a	CG126129-01	33	34	Epithelium differentiation factor (PEDF)
7b	CG126129-02	35	36	Epithelium differentiation factor (PEDF)
8a	CG142202-01	37	38	Cytokine Receptor CRL2 Precursor
8b	CG142202-03	39	40	Cytokine Receptor CRL2 Precursor
8c	CG142202-02	41	42	Cytokine Receptor CRL2 Precursor
9a	CG142621-01	43	44	Human GTP binding protein
10a	CG142761-01	45	46	Histocombatibility 13
11a	CG143926-01	47	48	HLA-B7 alpha chain precursor
12a	CG144193-01	49	50	Secreted phosphoprotein 24
12b	CG144193-02	51	52	Secreted phosphoprotein 24
13a	CG144545-01	53	54	Neuronal thread protein

14a	CG144884-01	55	56	B-Lymphocyte Activation Marker Blast-1 Precursor
14b	CG144884-02	57	58	B-Lymphocyte Activation Marker Blast-1 Precursor
15a	CG145122-01	59	60	MtN3/saliva Homolog
16a	CG145198-01	61	62	Secreted Protein
16b	278498076	63	64	Secreted Protein
16c	278498091	65	66	Secreted Protein
16d	CG145198-02	67	68	Secreted Protein
16e	CG145198-03	69	70	Secreted Protein
17a	CG145286-01	71	72	Tm6 protein
17b	CG145286-02	73	74	Tm6 protein
18a	CG145650-01	75	76	Lectin C-Type Domain Protein
18b	CG145650-02	77	78	Lectin C-Type Domain Protein
18c	CG145650-03	79	80	Lectin C-Type Domain Protein
19a	CG145836-01	81	82	STAR protein
19b	CG145836-02	83	84	STAR protein
20a	CG145978-01	85	86	DUF221 domain containing membrane protein
20b	CG145978-02	87	88	DUF221 domain containing membrane protein
21a	CG145997-01	89	90	Similar to Drosophila FRY gene protein
22a	CG146119-01	91	92	Papilin
23a	CG146202-01	93	94	Membrane-Associated Lectin Type-C
24a	CG146250-01	95	96	Membrane protein containing vwd domain
24b	CG146250-02	97	98	Membrane protein containing vwd domain
24c	CG146250-03	99	100	Membrane protein containing vwd domain
25a	CG146625-01	101	102	Type IIIa Membrane Protein
25b	CG146625-02	103	104	Type IIIa Membrane Protein
25c	CG146625-03	105	106	Type IIIa Membrane Protein
26a	CG147284-01	107	108	Cadherin 6
27a	CG147937-01	109	110	NK Cell Receptor CS-1
27b	CG147937-02	111	112	NK Cell Receptor CS-1
28a	CG148221-01	113	114	Claudin domain containing transmembrane protein
28b	CG148221-02	115	116	Claudin domain containing transmembrane protein
29a	CG148476-01	117	118	Membrane-bound protein PRO1383
30a	CG148818-01	119	120	Membrane protein KIAA0146
31a	CG149332-01	121	122	Interferon Induced Transmembrane Protein 3 (1-8U)
32a	CG149649-01	123	124	Type IIIA membrane protein
32b	CG149649-02	125	126	Type IIIA membrane protein
33a	CG149680-01	127	128	Pb39 (Prostate Cancer Overexpressed Gene 1)
33b	CG149680-02	129	130	Pb39 (Prostate Cancer Overexpressed Gene 1)
34a	CG149777-01	131	132	KIAA0575/Greb1
34b	CG149777-02	133	134	KIAA0575/Greb1
34c	257474374	135	136	Integrin Alpha-11

34d	257474386	137	138	Fibrillin
35a	CG150005-01	139	140	von Willebrand factor
36a	CG150189-01	141	142	von Willebrand factor
37a	CG150267-01	143	144	Fibrillin
38a	CG150362-01	145	146	Otoferlin
39a	CG150637-01	147	148	Cytokine Receptor
39b	CG150637-02	149	150	Cytokine Receptor
40a	CG150694-01	151	152	Cytokine Receptor
41a	CG151069-01	153	154	Bone marrow secreted protein
42a	CG151189-01	155	156	Human apoptosis protein (APOP-2)
43a	CG151801-01	157	158	Laminin 5-Beta 3
44a	CG165961-01	159	160	Fibulin 3
44b	CG165961-02	161	162	Fibulin 3
44c	CG165961-03	163	164	Fibulin 3
44d	CG165961-04	165	166	Epithelium differentiation factor (PEDF)
45a	CG171681-01	167	168	Cytokine Receptor CRL2 Precursor
45b	CG171681-02	169	170	Cytokine Receptor CRL2 Precursor
45c	CG171681-03	171	172	Cytokine Receptor CRL2 Precursor
46a	CG173318-01	173	174	Human metabolism protein 16
47a	CG51595-01	175	176	HLA-B7 alpha chain precursor
47b	CG51595-03	177	178	HLA-B7 alpha chain precursor
47c	CG51595-04	179	180	Secreted phosphoprotein 24
47d	CG51595-06	181	182	Secreted phosphoprotein 24
47e	CG51595-07	183	184	1700029J11RIK protein
47f	306395637	185	186	B-Lymphocyte Activation Marker Blast-1 Precursor
47g	CG51595-01	187	188	B-Lymphocyte Activation Marker Blast-1 Precursor
47h	283842727	189	190	MtN3/saliva Homolog
47i	283842704	191	192	MtN3/saliva Homolog
47j	CG51595-01	193	194	MtN3/saliva Homolog
47k	310658551	195	196	MtN3/saliva Homolog
47l	CG51595-02	197	198	MtN3/saliva Homolog
47m	CG51595-05	199	200	MtN3/saliva Homolog
48a	CG57209-01	201	202	Tm6 protein
48b	CG57209-03	203	204	Lectin C-Type Domain Protein
48c	CG57209-02	205	206	Lectin C-Type Domain Protein
48d	CG57209-04	207	208	Lectin C-Type Domain Protein
49a	CG57292-01	209	210	STAR protein
49b	CG57292-02	211	212	STAR protein
50a	CG97715-01	213	214	Transmembrane protein PT27

Table A indicates the homology of NOVX polypeptides to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table A will be useful

in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table A.

Pathologies, diseases, disorders and condition and the like that are associated with NOVX sequences include, but are not limited to: *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, metabolic disturbances associated with obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Ostoeodystrophy, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias,] the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers, as well as conditions such as transplantation, neuroprotection, fertility, or regeneration (*in vitro* and *in vivo*).

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table A, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of

small molecules that modulate or inhibit diseases associated with the protein families listed in Table A.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.* detection of a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOVX clones

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, *e.g.*, by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) a biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 107; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 107, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 107; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 107; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 107 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 107; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 107, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 107 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 107; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 107 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 107; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 107 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (*e.g.*, host

cell) in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probe", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), about 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single-stranded or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, *etc.*). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium, or of chemical precursors or other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, or a complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, is one that is sufficiently complementary to the nucleotide

sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

A "fragment" provided herein is defined as a sequence of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and is at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5' direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

A "derivative" is a nucleic acid sequence or amino acid sequence formed from the native compounds either directly, by modification or partial substitution. An "analog" is a nucleic acid sequence or amino acid sequence that has a structure similar to, but not identical to, the native compound, *e.g.* they differs from it in respect to certain components or side chains. Analogs may be synthetic or derived from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. A

"homolog" is a nucleic acid sequence or amino acid sequence of a particular gene that is derived from different species.

Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is

uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107; or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

“A polypeptide having a biologically-active portion of a NOVX polypeptide” refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described

below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding a NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from a human SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the

nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase “stringent hybridization conditions” refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM

EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, et *al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Krieger, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.,* Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* **78**: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be

introduced by mutation into the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, thereby leading to changes in the amino acid sequences of the encoded NOVX protein, without altering the functional ability of that NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 40% homologous to the amino acid sequences of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107; more preferably at least about 70% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107; still more preferably at least about 80% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107; even more preferably at least about 90% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107; and most preferably at least about 95% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, by standard techniques, such as site-directed mutagenesis and

PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of a nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (*i*) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (*ii*) complex formation between a mutant NOVX protein and a NOVX ligand; or (*iii*) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Interfering RNA

In one aspect of the invention, NOVX gene expression can be attenuated by RNA interference. One approach well-known in the art is short interfering RNA (siRNA) mediated gene silencing where expression products of a NOVX gene are targeted by specific double stranded NOVX derived siRNA nucleotide sequences that are complementary to at least a 19-25 nt long segment of the NOVX gene transcript, including the 5' untranslated (UT) region, the ORF, or the 3' UT region. *See*, *e.g.*, PCT applications WO00/44895, WO99/32619, WO01/75164, WO01/92513, WO 01/29058, WO01/89304, WO02/16620, and WO02/29858, each incorporated by reference herein in their entirety. Targeted genes can be a NOVX gene, or an upstream or downstream modulator of the NOVX gene. Nonlimiting examples of upstream or downstream modulators of a NOVX gene include, *e.g.*, a transcription factor that binds the NOVX gene promoter, a kinase or phosphatase that interacts with a NOVX polypeptide, and polypeptides involved in a NOVX regulatory pathway.

According to the methods of the present invention, NOVX gene expression is silenced using short interfering RNA. A NOVX polynucleotide according to the invention includes a siRNA polynucleotide. Such a NOVX siRNA can be obtained using a NOVX polynucleotide sequence, for example, by processing the NOVX ribopolynucleotide sequence in a cell-free system, such as but not limited to a Drosophila extract, or by transcription of recombinant double stranded NOVX RNA or by chemical synthesis of nucleotide sequences homologous to a NOVX sequence. *See*, *e.g.*, Tuschl, Zamore, Lehmann, Bartel and Sharp (1999), Genes & Dev. 13: 3191-3197, incorporated herein by reference in its entirety. When synthesized, a typical 0.2 micromolar-scale RNA synthesis provides about 1 milligram of siRNA, which is sufficient for 1000 transfection experiments using a 24-well tissue culture plate format.

The most efficient silencing is generally observed with siRNA duplexes composed of a 21-nt sense strand and a 21-nt antisense strand, paired in a manner to have a 2-nt 3' overhang. The sequence of the 2-nt 3' overhang makes an additional small contribution to the specificity of siRNA target recognition. The contribution to specificity is localized to the unpaired nucleotide adjacent to the first paired bases. In one embodiment, the nucleotides in the 3' overhang are ribonucleotides. In an alternative embodiment, the

nucleotides in the 3' overhang are deoxyribonucleotides. Using 2'-deoxyribonucleotides in the 3' overhangs is as efficient as using ribonucleotides, but deoxyribonucleotides are often cheaper to synthesize and are most likely more nuclease resistant.

A contemplated recombinant expression vector of the invention comprises a NOVX DNA molecule cloned into an expression vector comprising operatively-linked regulatory sequences flanking the NOVX sequence in a manner that allows for expression (by transcription of the DNA molecule) of both strands. An RNA molecule that is antisense to NOVX mRNA is transcribed by a first promoter (*e.g.*, a promoter sequence 3' of the cloned DNA) and an RNA molecule that is the sense strand for the NOVX mRNA is transcribed by a second promoter (*e.g.*, a promoter sequence 5' of the cloned DNA). The sense and antisense strands may hybridize in vivo to generate siRNA constructs for silencing of the NOVX gene. Alternatively, two constructs can be utilized to create the sense and anti-sense strands of a siRNA construct. Finally, cloned DNA can encode a construct having secondary structure, wherein a single transcript has both the sense and complementary antisense sequences from the target gene or genes. In an example of this embodiment, a hairpin RNAi product is homologous to all or a portion of the target gene. In another example, a hairpin RNAi product is a siRNA. The regulatory sequences flanking the NOVX sequence may be identical or may be different, such that their expression may be modulated independently, or in a temporal or spatial manner.

In a specific embodiment, siRNAs are transcribed intracellularly by cloning the NOVX gene templates into a vector containing, *e.g.*, a RNA pol III transcription unit from the smaller nuclear RNA (snRNA) U6 or the human RNase P RNA H1. One example of a vector system is the GeneSuppressor™ RNA Interference kit (commercially available from Imgenex). The U6 and H1 promoters are members of the type III class of Pol III promoters. The +1 nucleotide of the U6-like promoters is always guanosine, whereas the +1 for H1 promoters is adenosine. The termination signal for these promoters is defined by five consecutive thymidines. The transcript is typically cleaved after the second uridine. Cleavage at this position generates a 3' UU overhang in the expressed siRNA, which is similar to the 3' overhangs of synthetic siRNAs. Any sequence less than 400 nucleotides in length can be transcribed by these promoter, therefore they are ideally suited for the expression of around 21-nucleotide siRNAs in, *e.g.*, an approximately 50-nucleotide RNA stem-loop transcript.

A siRNA vector appears to have an advantage over synthetic siRNAs where long term knock-down of expression is desired. Cells transfected with a siRNA expression vector would experience steady, long-term mRNA inhibition. In contrast, cells transfected with exogenous synthetic siRNAs typically recover from mRNA suppression within seven days or ten rounds of cell division. The long-term gene silencing ability of siRNA expression vectors may provide for applications in gene therapy.

In general, siRNAs are chopped from longer dsRNA by an ATP-dependent ribonuclease called DICER. DICER is a member of the RNase III family of double-stranded RNA-specific endonucleases. The siRNAs assemble with cellular proteins into an endonuclease complex. *In vitro* studies in Drosophila suggest that the siRNAs/protein complex (siRNP) is then transferred to a second enzyme complex, called an RNA-induced silencing complex (RISC), which contains an endoribonuclease that is distinct from DICER. RISC uses the sequence encoded by the antisense siRNA strand to find and destroy mRNAs of complementary sequence. The siRNA thus acts as a guide, restricting the ribonuclease to cleave only mRNAs complementary to one of the two siRNA strands.

A NOVX mRNA region to be targeted by siRNA is generally selected from a desired NOVX sequence beginning 50 to100 nt downstream of the start codon. Alternatively, 5' or 3' UTRs and regions nearby the start codon can be used but are generally avoided, as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. An initial BLAST homology search for the selected siRNA sequence is done against an available nucleotide sequence library to ensure that only one gene is targeted. Specificity of target recognition by siRNA duplexes indicate that a single point mutation located in the paired region of an siRNA duplex is sufficient to abolish target mRNA degradation. See, Elbashir *et al.* 2001 EMBO J. 20(23):6877-88. Hence, consideration should be taken to accommodate SNPs, polymorphisms, allelic variants or species-specific variations when targeting a desired gene.

In one embodiment, a complete NOVX siRNA experiment includes the proper negative control. A negative control siRNA generally has the same nucleotide composition as the NOVX siRNA but lack significant sequence homology to the genome. Typically, one would scramble the nucleotide sequence of the NOVX siRNA and do a homology search to make sure it lacks homology to any other gene.

Two independent NOVX siRNA duplexes can be used to knock-down a target NOVX gene. This helps to control for specificity of the silencing effect. In addition, expression of two independent genes can be simultaneously knocked down by using equal concentrations of different NOVX siRNA duplexes, *e.g.*, a NOVX siRNA and an siRNA for a regulator of a NOVX gene or polypeptide. Availability of siRNA-associating proteins is believed to be more limiting than target mRNA accessibility.

A targeted NOVX region is typically a sequence of two adenines (AA) and two thymidines (TT) divided by a spacer region of nineteen (N19) residues (*e.g.*, AA(N19)TT). A desirable spacer region has a G/C-content of approximately 30% to 70%, and more preferably of about 50%. If the sequence AA(N19)TT is not present in the target sequence, an alternative target region would be AA(N21). The sequence of the NOVX sense siRNA corresponds to (N19)TT or N21, respectively. In the latter case, conversion of the 3' end of the sense siRNA to TT can be performed if such a sequence does not naturally occur in the NOVX polynucleotide. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. Symmetric 3' overhangs may help to ensure that the siRNPs are formed with approximately equal ratios of sense and antisense target RNA-cleaving siRNPs. *See*, *e.g.*, Elbashir, Lendeckel and Tuschl (2001). Genes & Dev. 15: 188-200, incorporated by reference herein in its entirely. The modification of the overhang of the sense sequence of the siRNA duplex is not expected to affect targeted mRNA recognition, as the antisense siRNA strand guides target recognition.

Alternatively, if the NOVX target mRNA does not contain a suitable AA(N21) sequence, one may search for the sequence NA(N21). Further, the sequence of the sense strand and antisense strand may still be synthesized as 5' (N19)TT, as it is believed that the sequence of the 3'-most nucleotide of the antisense siRNA does not contribute to specificity. Unlike antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing. *See*, Harborth, *et al.* (2001) J. Cell Science 114: 4557-4565, incorporated by reference in its entirety.

Transfection of NOVX siRNA duplexes can be achieved using standard nucleic acid transfection methods, for example, OLIGOFECTAMINE Reagent (commercially available from Invitrogen). An assay for NOVX gene silencing is generally performed approximately 2 days after transfection. No NOVX gene silencing has been observed in the absence of transfection reagent, allowing for a comparative analysis of the wild-type

and silenced NOVX phenotypes. In a specific embodiment, for one well of a 24-well plate, approximately 0.84 μg of the siRNA duplex is generally sufficient. Cells are typically seeded the previous day, and are transfected at about 50% confluence. The choice of cell culture media and conditions are routine to those of skill in the art, and will vary with the choice of cell type. The efficiency of transfection may depend on the cell type, but also on the passage number and the confluency of the cells. The time and the manner of formation of siRNA-liposome complexes (*e.g.* inversion versus vortexing) are also critical. Low transfection efficiencies are the most frequent cause of unsuccessful NOVX silencing. The efficiency of transfection needs to be carefully examined for each new cell line to be used. Preferred cell are derived from a mammal, more preferably from a rodent such as a rat or mouse, and most preferably from a human. Where used for therapeutic treatment, the cells are preferentially autologous, although non-autologous cell sources are also contemplated as within the scope of the present invention.

For a control experiment, transfection of 0.84 μg single-stranded sense NOVX siRNA will have no effect on NOVX silencing, and 0.84 μg antisense siRNA has a weak silencing effect when compared to 0.84 μg of duplex siRNAs. Control experiments again allow for a comparative analysis of the wild-type and silenced NOVX phenotypes. To control for transfection efficiency, targeting of common proteins is typically performed, for example targeting of lamin A/C or transfection of a CMV-driven EGFP-expression plasmid (*e.g.* commercially available from Clontech). In the above example, a determination of the fraction of lamin A/C knockdown in cells is determined the next day by such techniques as immunofluorescence, Western blot, Northern blot or other similar assays for protein expression or gene expression. Lamin A/C monoclonal antibodies may be obtained from Santa Cruz Biotechnology.

Depending on the abundance and the half life (or turnover) of the targeted NOVX polynucleotide in a cell, a knock-down phenotype may become apparent after 1 to 3 days, or even later. In cases where no NOVX knock-down phenotype is observed, depletion of the NOVX polynucleotide may be observed by immunofluorescence or Western blotting. If the NOVX polynucleotide is still abundant after 3 days, cells need to be split and transferred to a fresh 24-well plate for re-transfection. If no knock-down of the targeted protein is observed, it may be desirable to analyze whether the target mRNA (NOVX or a NOVX upstream or downstream gene) was effectively destroyed by the transfected siRNA duplex. Two days after transfection, total RNA is prepared, reverse transcribed using a

target-specific primer, and PCR-amplified with a primer pair covering at least one exon-exon junction in order to control for amplification of pre-mRNAs. RT/PCR of a non-targeted mRNA is also needed as control. Effective depletion of the mRNA yet undetectable reduction of target protein may indicate that a large reservoir of stable NOVX protein may exist in the cell. Multiple transfection in sufficiently long intervals may be necessary until the target protein is finally depleted to a point where a phenotype may become apparent. If multiple transfection steps are required, cells are split 2 to 3 days after transfection. The cells may be transfected immediately after splitting.

An inventive therapeutic method of the invention contemplates administering a NOVX siRNA construct as therapy to compensate for increased or aberrant NOVX expression or activity. The NOVX ribopolynucleotide is obtained and processed into siRNA fragments, or a NOVX siRNA is synthesized, as described above. The NOVX siRNA is administered to cells or tissues using known nucleic acid transfection techniques, as described above. A NOVX siRNA specific for a NOVX gene will decrease or knockdown NOVX transcription products, which will lead to reduced NOVX polypeptide production, resulting in reduced NOVX polypeptide activity in the cells or tissues.

The present invention also encompasses a method of treating a disease or condition associated with the presence of a NOVX protein in an individual comprising administering to the individual an RNAi construct that targets the mRNA of the protein (the mRNA that encodes the protein) for degradation. A specific RNAi construct includes a siRNA or a double stranded gene transcript that is processed into siRNAs. Upon treatment, the target protein is not produced or is not produced to the extent it would be in the absence of the treatment.

Where the NOVX gene function is not correlated with a known phenotype, a control sample of cells or tissues from healthy individuals provides a reference standard for determining NOVX expression levels. Expression levels are detected using the assays described, *e.g.*, RT-PCR, Northern blotting, Western blotting, ELISA, and the like. A subject sample of cells or tissues is taken from a mammal, preferably a human subject, suffering from a disease state. The NOVX ribopolynucleotide is used to produce siRNA constructs, that are specific for the NOVX gene product. These cells or tissues are treated by administering NOVX siRNA's to the cells or tissues by methods described for the transfection of nucleic acids into a cell or tissue, and a change in NOVX polypeptide or polynucleotide expression is observed in the subject sample relative to the control sample,

using the assays described. This NOVX gene knockdown approach provides a rapid method for determination of a NOVX minus (NOVX⁻) phenotype in the treated subject sample. The NOVX⁻ phenotype observed in the treated subject sample thus serves as a marker for monitoring the course of a disease state during treatment.

In specific embodiments, a NOVX siRNA is used in therapy. Methods for the generation and use of a NOVX siRNA are known to those skilled in the art. Example techniques are provided below.

Production of RNAs

Sense RNA (ssRNA) and antisense RNA (asRNA) of NOVX are produced using known methods such as transcription in RNA expression vectors. In the initial experiments, the sense and antisense RNA are about 500 bases in length each. The produced ssRNA and asRNA (0.5 μM) in 10 mM Tris-HCl (pH 7.5) with 20 mM NaCl were heated to 95° C for 1 min then cooled and annealed at room temperature for 12 to 16 h. The RNAs are precipitated and resuspended in lysis buffer (below). To monitor annealing, RNAs are electrophoresed in a 2% agarose gel in TBE buffer and stained with ethidium bromide. See, *e.g.*, Sambrook et al., Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, N.Y. (1989).

Lysate Preparation

Untreated rabbit reticulocyte lysate (Ambion) are assembled according to the manufacturer's directions. dsRNA is incubated in the lysate at 30° C for 10 min prior to the addition of mRNAs. Then NOVX mRNAs are added and the incubation continued for an additional 60 min. The molar ratio of double stranded RNA and mRNA is about 200:1. The NOVX mRNA is radiolabeled (using known techniques) and its stability is monitored by gel electrophoresis.

In a parallel experiment made with the same conditions, the double stranded RNA is internally radiolabeled with a ^{32}P-ATP. Reactions are stopped by the addition of 2 X proteinase K buffer and deproteinized as described previously (Tuschl *et al.*, Genes Dev., 13:3191-3197 (1999)). Products are analyzed by electrophoresis in 15% or 18% polyacrylamide sequencing gels using appropriate RNA standards. By monitoring the gels for radioactivity, the natural production of 10 to 25 nt RNAs from the double stranded RNA can be determined.

The band of double stranded RNA, about 21-23 bps, is eluded. The efficacy of these 21-23 mers for suppressing NOVX transcription is assayed in vitro using the same rabbit reticulocyte assay described above using 50 nanomolar of double stranded 21-23 mer for each assay. The sequence of these 21-23 mers is then determined using standard nucleic acid sequencing techniques.

RNA Preparation

21 nt RNAs, based on the sequence determined above, are chemically synthesized using Expedite RNA phosphoramidites and thymidine phosphoramidite (Proligo, Germany). Synthetic oligonucleotides are deprotected and gel-purified (Elbashir, Lendeckel, & Tuschl, Genes & Dev. 15, 188-200 (2001)), followed by Sep-Pak C18 cartridge (Waters, Milford, Mass., USA) purification (Tuschl, et al., Biochemistry, 32:11658-11668 (1993)).

These RNAs (20 μM) single strands are incubated in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) for 1 min at 90° C followed by 1 h at 37° C.

Cell Culture

A cell culture known in the art to regularly express NOVX is propagated using standard conditions. 24 hours before transfection, at approx. 80% confluency, the cells are trypsinized and diluted 1:5 with fresh medium without antibiotics (1-3 X 105 cells/ml) and transferred to 24-well plates (500 ml/well). Transfection is performed using a commercially available lipofection kit and NOVX expression is monitored using standard techniques with positive and negative control. A positive control is cells that naturally express NOVX while a negative control is cells that do not express NOVX. Base-paired 21 and 22 nt siRNAs with overhanging 3' ends mediate efficient sequence-specific mRNA degradation in lysates and in cell culture. Different concentrations of siRNAs are used. An efficient concentration for suppression in vitro in mammalian culture is between 25 nM to 100 nM final concentration. This indicates that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or ribozyme gene targeting experiments.

The above method provides a way both for the deduction of NOVX siRNA sequence and the use of such siRNA for in vitro suppression. In vivo suppression may be

performed using the same siRNA using well known in vivo transfection or gene therapy transfection techniques.

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using

chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 5-methoxyuracil, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 2-thiouracil, 4-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense

molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330.

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease

activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleotide bases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomer can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*See,* Hyrup, *et al.,* 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*See,* Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.,* 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow

DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation (*see*, Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, a NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one

embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107, and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then

compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.,* A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.,* the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not

substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically-active portions of a NOVX protein. In yet another embodiment, a NOVX fusion protein comprises at least three biologically-active portions of a NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is a NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.,* mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is

generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of

vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. *See, e.g.*, Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 107, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes

encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. A NOVX polypeptide or a fragment thereof comprises at least one antigenic epitope. An anti-NOVX antibody of the present invention is said to specifically bind to antigen NOVX when the equilibrium binding constant (K_D) is $\leq 1\ \mu M$, preferably ≤ 100 nM, more preferably ≤ 10 nM, and most preferably ≤ 100 pM to about 1 pM, as measured by assays such as radioligand binding assays or similar assays known to those skilled in the art.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components..

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, *etc.*), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular

species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J.

Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding,1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13

(1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies"

herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the

Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a

protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers

which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (*e.g.* tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.* alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e.g.* $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were

reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (*e.g.* CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include

iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as

dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is in turn conjugated to a cytotoxic agent.

Immunoliposomes

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al ., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon *et al.*, J. National Cancer Inst., 81(19): 1484 (1989).

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme linked immunosorbent assay (ELISA)

and other immunologically mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Antibodies directed against a NOVX protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies specific to a NOVX protein, or derivative, fragment, analog or homolog thereof, that contain the antibody derived antigen binding domain, are utilized as pharmacologically active compounds (referred to hereinafter as "Therapeutics").

An antibody specific for a NOVX protein of the invention (e.g., a monoclonal antibody or a polyclonal antibody) can be used to isolate a NOVX polypeptide by standard techniques, such as immunoaffinity, chromatography or immunoprecipitation. An antibody to a NOVX polypeptide can facilitate the purification of a natural NOVX antigen from cells, or of a recombinantly produced NOVX antigen expressed in host cells. Moreover, such an anti-NOVX antibody can be used to detect the antigenic NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic NOVX protein. Antibodies directed against a NOVX protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of

bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Therapeutics

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, *e.g.*, Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting

immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulus, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Thory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-an analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector,

"operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, *etc.*).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the

fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly

used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a

discussion of the regulation of gene expression using antisense genes *see, e.g.*, Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection

(*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, *etc.* A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant

female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby

alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See*, O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic

animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for

parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required

other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable,

biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a

NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra.*

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound

form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, *etc.*), detecting catalytic/enzymatic

activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises

determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target

molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also involved in the propagation of signals by

the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a

chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.*, D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600

bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for

RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 107, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an

individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:$2n$-1, wherein n is an integer between 1 and 107, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or

500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and

comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic

acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from a NOVX gene; (*ii*) an addition of one or more nucleotides to a NOVX gene; (*iii*) a substitution of one or more nucleotides of a NOVX gene, (*iv*) a chromosomal rearrangement of a NOVX gene; (*v*) an alteration in the level of a messenger RNA transcript of a NOVX gene, (*vi*) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (*viii*) a non-wild-type level of a NOVX protein, (*ix*) allelic loss of a NOVX gene, and (*x*) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or

detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al., supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.*, Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.*, Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.*, Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662.

According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.*, Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.*, Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.*, Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.*, Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.*, Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele

specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.*, Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.*, Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.*, Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.*, Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly

polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule)

that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those

diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

These methods of treatment will be discussed more fully, below.

Diseases and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see*, *e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, *etc.*) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another

embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situ*ations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from diseases, disorders, conditions and the like, including but not limited to those listed herein.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein

the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example A: Polynucleotide and Polypeptide Sequences, and Homology Data

Example 1.

The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 1A.

Table 1A. NOV1 Sequence Analysis			
	SEQ ID NO: 1	8482 bp	
NOV1a, CG105472-01 DNA Sequence	AAATAAAGTTTTTTCAATGGAAGGCTTGCAGCTCTTGAGGACCTGCCAAATGGAAGAAGGACAGAGAC CTGGAGCCCTATGGAAAGTTCTGACACCATGTGTGGAAGGACATGGCTTTTAACACGTGTGGTGACTG GAGTAGCTGCAGCTGAGGACAGCCACCCTTTCTTCGTCTCTGCTGAGCGAAGGCTACACGGCCCTTCC TCCTTGCAGCTGTTTCACCTTCTACCTTGCGTGGAGCCAGGCTTTTGCACCGAATCTGAGATGCCATT TTAAACAGAAGACTCCATCCTCTTGAAG**ATG**GGAAATTCTTACGCTGGACAGCTGAAGACGACACGCT TTGAAGAGGTCTTGCACAATTCCATCGAGGCATCCCTGCGGTCCAACAACCTGGTGCCCAGGCCCATC TTTTCCCAGCTGTACCTGGAAGCTGAGCAGCAGCTTGCCGCTCTAGAAGGTGGTAGCCGAGTGGACAA TGAGGAAGAGGAAGAAGAGGGAGAAGGAGGGCTGGAAACAAATGGCCCCCCAAACCCTTTCCAGCTGC ACCCTCTGCCTGAAGGATGCTGTACCACAGACGGTTTTTGCCAGGCCGGGAAGGACCTGCGCCTTGTC TCCATTTCCAACGAGCCCATGGATGTCCCTGCGGGCTTTCTCCTCGTGGGGGTCAAGTCCCCCAGCCT GCCGGACCATCTCCTGGTGTGCGCCGTTGACAAGAGGTTCTTGCCAGATGACAATGGCCACAATGCTC TTCTTGGTTTCTCTGGGAATTGTGTTGGCTGTGGAAAGAAAGGCTTCTGTTACTTCACGGAATTCTCC AATCATATAAATCTGAAACTGACCACTCAACCCAAGAAGCAGAAACACTTGAAGTATTACCTGGTCCG TAATGCACAAGGGACTCTAACCAAAGGACCTTTAATCTGTTGGAAAGGCTCAGAGTTTAGAAGCCGGC AGATCCCCGCCAGTACTTGTTCCAGTTCCCTCTTCCCAGCCCTGGAGAGCACGGCTGCCTTCCCCAGC GAGCCCGTTCCTGGGACGAACCCCAGCATCCTGATGGGAGCTCAGCAGGCAGGTCCAGCTTCTGATCA CCCCTCACTAAACGCAGCAATGGGTCCGGCTGTTTTCAACGGCAAAGATTCCCCGAAGTGCCAACAAC TGGCAAAGAATAACCTGTTGGCCCTGCCGCGACCATCGGCTTTAGGTATCTTGTCAAACTCCGGGCCC CCCAAAAAACGCCACAAAGGGTGGTCTCCAGAATCTCCATCAGCTCCAGATGGTGGCTGCCCCCAAGG TGGTGGGAACAGAGCTAAGTATGAGAGCGCAGGCATGTCCTGCGTGCCGCAGGTTGGCTTGGTGGGAC CAGCTTCAGTCACCTTTCCAGTGGTGGCCTCTGGAGAACCAGTGTCTGTTCCTGACAACTTGCTGAAA ATATGCAAGGCCAAGCCAGTGATATTTAAAGGCGATGGGAACTTCCCTTACCTCTGTGGGAACCTGAA TGACGTCGTGGTCAGCCCCCTCTTGTACACGTGCTACCAGAATTCCCAGTCTGTCTCACGGGCATACG AGCAGTACGGCGCCTCTGCCATCCAGCCCATCTCCGAGGAGATGCAGCTCCTGCTTACCGTCTACTAC CTGGTCCAGCTGGCCGCGGACCAGGTGCCCTTGATGGAGGACCTGGAGCAGATCTTCCTGCGCTCTTG GCGCGAGTCGCACCTGACCGAGATCCGGCAGTACCAGCAGGCGCCGCCGCAGCCCTTCCCGCCCGCGC CCAGCGCCGCGGCACCCGTGACCTCCGCGCAGCTGCCCTGGCTGGCCAGCCTGGCCGCCAGCTCCTGC AACGACAGCGTGCACGTCATCGAGTGTGCTTACTCCCTGGCCGAGGGCCTCTCCGAGATGTTCCGGCT GTTGGTCGAGGGCAAGCTTGCCAAGACCAACTACGTGGTCATCATCTGCGCCTGCCGCAGCGCGGCCA TCGACTCCTGCATCGCCGTCACCGGTAAATACCAAGCCCGGATTCTTTCCGAGAGCCTTCTCACTCCT GCGGAGTACCAGAAGGAAGTCAATTACGAGCTGGTTACGGGGAAGGTAGACTCGCTGGGGGCCTTCTT TAGCACCCTCTGTCCAGAGGGTGACATTGACATTTTGCTGGACAAATTTCACCAGGAAAATCAAGGCC ATATTTCTTCCTCACTCGCTGCCTCTTCTGTCACTAAAGCAGCATCCCTGGATGTCAGTGGGACACCG GTGTGCACAAGTTACAATCTGGAGCCACACAGCATCCGGCCCTTCCAGCTGGCAGTAGCGCAGAAGCT CCTCTCCCATGTGTGTTCCATTGCGGATTCCAGCACCCAAAATCTGGACCTGGGATCCTTTGAGAAGG TGGACTTTCTCATTTGCATTCCCCCCTCAGAAGTGACCTACCAGCAGACTCTGCTCCATGTGTGGCAT TCAGGTGTTTTGCTGGAGCTTGGTCTGAAGAAAGAGCACATGACGAAGCAGAGGGTGGAACAGTATGT TCTGAAGCTAGACACGGAGGCACAGACAAAATTTAAGGCTTTTCTGCAAAACTCCTTCCAGAACCCGC ATACACTTTTTGTCCTAATCCATGACCATGCGCACTGGGATCTTGTGAGTAGCACTGTTCATAACCTC		

	TATTCTCAAAGTGACCCGTCGGTGGGATTGGTGGACCGATTGCTCAACTGCAGGGAGGTGAAGGAGGC
	CCCCAACATTGTGACACTTCACGTGACCTCCTTCCCGTATGCACTGCAGACACAGCACACCCTCATCA
	GCCCCTACAACGAGATCCACTGGCCTGCCTCCTGCAGTAATGGAGTGGACTTATATCATGAAAATAAG
	AAGTACTTCGGGCTGTCGGAGTTTATTGAATCCACCCTTTCAGGACACAGCCTCCCCTTGCTCAGATA
	CGATAGCTCCTTTGAGGCCATGGTCACTGCATTAGGAAAAAGGTTCCCCCGCCTGCACAGCGCGGTGA
	TCAGGACCTTTGTTCTCGTGCAGCACTACGCGGCCGCCCTGATGGCCGTAAGCGGCCTCCCGCAGATG
	AAGAACTACACGTCGGTGGAGACGCTGGAGATCACGCAGAACCTCCTCAACTCCCCGAAGCAGTGCCC
	CTGCGGCCACGGGCTCATGGTCCTGCTGCGGGTGCCCTGTTCGCCCCTGGCGGTGGTGGCCTATGAGC
	GGCTGGCCCACGTGCGGGCCCGGCTGGCGCTGGAGGAGCACTTTGAGATCATCCTGGGCAGTCCCAGC
	TCAGGCGTCACCGTGGGGAAGCACTTCGTAAAGCAGCTCAGGGTATGGCAGAAAATTGAGGATGTGGA
	GTGGAGACCCCAGACTTACTTGGAGCTGGAGGGTCTGCCTTGCATCCTGATCTTCAGTGGGATGGACC
	CGCATGGGGAGTCCTTGCCGAGGTCTTTGAGGTACTGTGACCTGCGATTGATAAACTCCTCCTGCTTG
	GTGAGAACAGCCTTGGAGCAGGAGCTGGGCCTGGCTGCCTACTTTGTGAGCAACGAGGTTCCCTTGGA
	GAAGGGGGCTAGGAACGAGGCCTTGGAGAGTGATGCTGAGAAGCTGAGCAGCACAGACAACGAGGATG
	AGGAGCTGGGGACAGAAGGCTCTACCTCGGAGAAGAGAAGCCCCATGAAAAGGGAGAGGTCCCGCTCC
	CACGACTCAGCATCCTCATCCCTCTCCTCCAAGGCTTCCGGTTCCGCGCTCGGTGGCGAGTCCTCGGC
	TCAGCCCACAGCACTCCCCCAGGGAGAGCATGCCAGGTCGCCCCAGCCCCGTGGCCCCGCAGAGGAGG
	GCAGAGCCCCTGGTGAGAAACAGAGGCCCCGGGCAAGTCAGGGGCCACCCTCGGCCATCAGCAGGCAC
	AGTCCCGGGCCGACGCCCCAGCCCGACTGTAGCCTCAGGACCGGCCAGAGGAGCGTCCAGGTGTCGGT
	CACCTCGTCGTGCTCCCAGCTGTCCTCCTCCTCGGGCTCATCCTCCTCATCCGTGGCGCCCGCTGCCG
	GCACGTGGGTCCTGCAGGCCTCCCAGTGCTCCTTGACCAAGGCCTGCCGCCAGCCACCCATTGTCTTC
	TTGCCCAAGCTCGTGTACGACATGGTTGTGTCCACTGACAGCAGTGGCCTGCCCAAGGCCGCCTCCCT
	CCTGCCCTCCCCCTCGGTCATGTGGGCCAGCTCTTTCCGCCCCCTGCTCAGCAAGACCATGACATCCA
	CCGAGCAGTCCCTCTACTACCGGCAGTGGACGGTGCCCCGGCCCAGCCACATGGACTACGGCAACCGG
	GCCGAGGGCCGCGTGGACGGCTTCCACCCCCGCAGGCTGCTGCTCAGCGGCCCCCCTCAGATCGGGAA
	GACAGGTGCCTACCTGCAGTTCCTCAGTGTCCTGTCCAGGATGCTTGTTCGGCTCACAGAAGTGGATG
	TCTATGACGAGGAGGAGATCAATATCAACCTGAGAGAAGAATCTGACTGGCATTATCTCCAGCTTAGC
	GACCCCTGGCCAGACCTGGAGCTGTTCAAGAAGTTGCCCTTTGACTACATCATTCACGACCCGAAGTA
	TGAAGATGCCAGCCTGATTTGTTCGCACTATCAGGGTATAAAGAGTGAAGACAGAGGGATGTCCCGGA
	AGCCGGAGGACCTTTATGTGCGGCGTCAGACGGCACGGATGAGACTGTCCAAGTACGCAGCGTACAAC
	ACTTACCACCACTGTGAGCAGTGCCACCAGTACATGGGCTTCCACCCCCGCTACCAGCTGTATGAGTC
	CACCCTGCACGCCTTTGCCTTCTCTTACTCCATGCTAGGAGAGGAGATCCAGCTGCACTTCATCATCC
	CCAAGTCCAAGGAGCACCACTTTGTCTTCAGCCAACCTGGAGGCCAGCTGGAGAGCATGCGACTACCC
	CTCGTGACAGACAAGAGCCATGAATATATAAAAAGTCCGACATTCACTCCAACCACCGGCCGTCACGA
	ACATGGGCTCTTTAATCTGTACCACGCAATGGACGGTGCCAGCCATTTGCACGTGCTGGTTGTCAAGG
	AATACGAGATGGCAATTTATAAGAAATATTGGCCCAACCACATCATGCTGGTGCTCCCCAGTATCTTC
	AACAGTGCTGGAGTTGGTGCTGCTCATTTCCTCATCAAGGAGCTGTCCTACCATAACCTGGAGCTCGA
	GCGGAACCGGCAGGAGGAGCTGGGAATCAAGCCGCAGGACATCTGGCCTTTCATTGTGATCTCTGATG
	ACTCCTGCGTGATGTGGAACGTGGTGGATGTCAACTCTGCTGGGGAGAGAAGCAGGGAGTTCTCCTGG
	TCGGAAAGGAACGTGTCTTTGAAGCACATCATGCAGCACATCGAGGCGGCCCCCGACATCATGCACTA
	CGCCCTGCTGGGCCTGCGGAAGTGGTCCAGCAAGACCCGGGCCAGCGAGGTGCAAGAGCCCTTCTCCC
	GCTGCCACGTGCACAACTTCATCATCCTGAACGTGGACCTGACCCAGAACGTGCAGTACAACCAGAAC
	CGGTTCCTGTGTGACGATGTAGACTTCAACCTGCGGGTGCACAGCGCCGGCCTCCTGCTCTGCCGGTT
	CAACCGCTTCAGCGTGATGAAGAAGCAGATCGTGGTGGGCGGCCACAGGTCCTTCCACATCACATCCA
	AGGTGTCTGATAACTCTGCCGCGGTCGTGCCGGCCCAGTACATCTGTGCCCCGGACAGCAAGCACACG
	TTCCTCGCAGCGCCCGCCCAGCTCCTGCTGGAGAAGTTCCTGCAGCACCACAGCCACCTCTTCTTCCC
	GCTGTCCCTGAAGAACCATGACCACCCAGTGCTGTCTGTCGACTGTTACCTGAACCTGGGATCTCAGA
	TTTCTGTTTGCTATGTGAGCTCCAGGCCCCACTCTTTAAACATCAGCTGCTCGGACTTGCTGTTCAGT
	GGGCTGCTGCTGTACCTCTGTGACTCTTTTGTGGGAGCTAGCTTTTTGAAAAAGTTTCATTTTCTGAA
	AGGTGCGACGTTGTGTGTCATCTGTCAGGACCGGAGCTCACTGCGCCAGACGGTCGTCCGCCTGGAGC
	TCGAGGACGAGTGGCAGTTCCGGCTGCGCGATGAGTTCCAGACCGCCAATGCCAGGGAAGACCGGCCG
	CTCTTTTTTCTGACGGACGACACATC**TGA**GGAAGACAGCGGCGAGTTTTCTGAAGAGATGAGTGCTC
	AGAGCCCTCATGCTGTTGAGGCTAAAGGGAGGCCTGGAACGGTGGGGCGTTTGACTGGAATGGACCCC
	AGGGACTGTCCAGGTGCAGCCCCTCCTAGTACACATGGGCCCCCGAGGCCGTGGTCCTGGGAGCCAGG
	AAGACTCCGCAGTGGGTGAGAATGAAAACTTGAGACTCCCAAGTTCTGGGCCAGCCCATTGCTCTGGG
	CTGTTTTAAAGCCCATTTCACGAGGAACAAAGATTTACTTCCTGTCCTGCCATTCGTGTGCTTCCATG
	GACAAACCTGATTTTTTTCTCTTAGTTCTAAAGAATCTTGGGTTATTTTGTAGCGGTGCCAGTATTTC
	AGTAGATGGGATTTCAGCCAAGTAGGTTCCCCTGTAACCTCCTACAAAGCAATATTCCAAAGGAACAT
	TTTAACTGTAAAGGCTGGAGACAAGAAAAAATAAGTAGATCGTTTTAATAACAATTATTTAATTGCCT
	ATAAGTTTGCTGTTTCAGAGGCTAGCCCAAAGGCATCAAATTTAATAAAGTTAAACAAATTGATTTAC
	TTCAGAGCAAATATGATCCTATTAAAATAATATAGGGTAAATACCCTACCTCTTAGAAAGGGCAAAAA
	TGCAAAGAAGCTTTCTTTAAAACTAAAAGGGTTTTTTGGGGGGGAGTTGGCGGGGAGGAAATAAGGC
	TAACAGAGGTTGACCTAAAATTAGCCTTACAAAGGAGAAAGGACCACATTGCTTACTTGAAACAGACA
	ATGAAAACAACCAAAGTGATATATAAAATAGTTGATGAGAACTAGACTTATGACTGTAGTTTACTAGA
	GTTTAGTTTTCAGTTGCTGAAGTAGCTCATTTTCTCTTACTAATGTTTGGTTCCTCAGGGAAGAATCT
	CACTTGACTAGAGAGGAGGTGGGAACAGAAGAGAGAAGGAGGCAGGGAGATGTATTTCTTAGGGCTCA
	CCCCTTCACAGACTGACAGAATGGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTTGAGATGGAC
	TCTAGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCGG
	GTTCTCACCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCACCACCACGCCCGG
	CTAATTTTTTGTATTTTTTAGTAGAGACGGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTCCTG
	ACCTCGTGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGTGCCTGCC
	CCAGAATGGTTTTTAAAGCCACAGTTGAGAGGCCACCCATTGCCCGGCGCCTGGACAGTGATCATCTT
	GTTCATCTTGTTCAGTCCTTTCTTGTGTGATTGGAATTATTCATCCCCTTTGAAAGATGAGAAGGTTG
	AGATGCAAAGAGTCTACCTTTCCAAGTTCTCACTGCTGGAAAGAGCTAGAAGCACAGTTCAAAGTTCT
	GGCTTCTGGACTCTGCAGTCCAGGTCTCCCTTCTCCCACTTGCCTACCCTCAATGCCACACTGTTTTT
	GAAGTGGCCCATAACTTGAAGGAAAAGTTTAAAGACAGTTCAATTTAATCATCAGAATGCATTCTTTT

	TTTTTTCGGAGACGGAGTTTCACTCTTGCTGCCCAGGCTGGAGTGCAATGGTGCAATGATCTCGGCTC ACTGCAACCTCTGCCTCCTGGGTTCAAGTGATTCTCCAGCCTCAGCCTCCCGAGTAGCTGGGATTATG GGCGCCCACCACCATGCCCAGCTAATTTTTGTATTTTTTTTTTTTAGTAGAGATGGGGTTTCGCCAGG TTGGCCAGGCTGGTCTTGTGAACTCCTGGCCTCAGGTGATCTGCCCACCTCATCCTCCAAAAGTGCTG GGATTACAGGCATGAGCCACTGCGCCTGGCCTCAGAATGCATTCTTACACATCTATCCTAGACATTTA TAAGCACTCTAATGGATAACAATCCAAGAATAAATGATTGTAAAAGATGATGCCGAAGAGTTGATGTC AATCTTTTTTTCCTAAGAAAAAAAGTCCGCGAGTATTAAATATTTAGATCAATGTTTATAAAATGATT ACTTTGTATATCTCATTATTCCTATTTTGGAATAAAAACTGACCTTCTTTAATCATATACTTGTCTTT TGTAAATAGCAGCTTTTGTGTCATTCTCCCCACTTTATTAGTTAATTTAAATTGGAAAAAACCCTCAA ACTAATATTCTTGTCTGTTCCAGTCTTATAAATAAAACTTATAATGCATG		
	ORF Start: ATG at 301		ORF Stop: TGA at 6148

	SEQ ID NO: 2	1949 aa	MW at 216410.6kD
NOV1a, CG105472-01 Protein Sequence	MGNSYAGQLKTTRFEEVLHNSIEASLRSNNLVPRPIFSQLYLEAEQQLAALEGGSRVDNEEEEEEGEG GLETNGPPNPFQLHPLPEGCCTTDGFCQAGKDLRLVSISNEPMDVPAGFLLVGVKSPSLPDHLLVCAV DKRFLPDDNGHNALLGFSGNCVGCGKKGFCYFTEFSNHINLKLTTQPKKQKHLKYYLVRNAQGTLTKG PLICWKGSEFRSRQIPASTCSSSLFPALESTAAFPSEPVPGTNPSILMGAQQAGPASDHPSLNAAMGP AVFNGKDSPKCQQLAKNNLLALPRPSALGILSNSGPPKKRHKGWSPESPSAPDGGCPQGGGNRAKYES AGMSCVPQVGLVGPASVTFPVVASGEPVSVPDNLLKICKAKPVIFKGDGNFPYLCGNLNDVVVSPLLY TCYQNSQSVSRAYEQYGASAIQPISEEMQLLLTVYYLVQLAADQVPLMEDLEQIFLRSWRESHLTEIR QYQQAPPQPFPPAPSAAAPVTSAQLPWLASLAASSCNDSVHVIECAYSLAEGLSEMFRLLVEGKLAKT NYVVIICACRSAAIDSCIAVTGKYQARILSESLLTPAEYQKEVNYELVTGKVDSLGAFFSTLCPEGDI DILLDKFHQENQGHISSSLAASSVTKAASLDVSGTPVCTSYNLEPHSIRPFQLAVAQKLLSHVCSIAD SSTQNLDLGSFEKVDFLICIPPSEVTYQQTLLHVWHSGVLLELGLKKEHMTKQRVEQYVLKLDTEAQT KFKAFLQNSFQNPHTLFVLIHDHAHWDLVSSTVHNLYSQSDPSVGLVDRLLNCREVKEAPNIVTLHVT SFPYALQTQHTLISPYNEIHWPASCSNGVDLYHENKKYFGLSEFIESTLSGHSLPLLRYDSSFEAMVT ALGKRFPRLHSAVIRTFVLVQHYAAALMAVSGLPQMKNYTSVETLEITQNLLNSPKQCPCGHGLMVLL RVPCSPLAVVAYERLAHVRARLALEEHFEIILGSPSSGVTVGKHFVKQLRVWQKIEDVEWRPQTYLEL EGLPCILIFSGMDPHGESLPRSLRYCDLRLINSSCLVRTALEQELGLAAYFVSNEVPLEKGARNEALE SDAEKLSSTDNEDEELGTEGSTSEKRSPMKRERSRSHDSASSSLSSKASGSALGGESSAQPTALPQGE HARSPQPRGPAEEGRAPGEKQRPRASQGPPSAISRHSPGPTPQPDCSLRTGQRSVQVSVTSSCSQLSS SSGSSSSSVAPAAGTWVLQASQCSLTKACRQPPIVFLPKLVYDMVVSTDSSGLPKAASLLPSPSVMWA SSFRPLLSKTMTSTEQSLYYRQWTVPRPSHMDYGNRAEGRVDGFHPRRLLLSGPPQIGKTGAYLQFLS VLSRMLVRLTEVDVYDEEEININLREESDWHYLQLSDPWPDLELFKKLPFDYIIHDPKYEDASLICSH YQGIKSEDRGMSRKPEDLYVRRQTARMRLSKYAAYNTYHHCEQCHQYMGFHPRYQLYESTLHAFAFSY SMLGEEIQLHFIIPKSKEHHFVFSQPGGQLESMRLPLVTDKSHEYIKSPTFTPTTGRHEHGLFNLYHA MDGASHLHVLVVKEYEMAIYKKYWPNHIMLVLPSIFNSAGVGAAHFLIKELSYHNLELERNRQEELGI KPQDIWPFIVISDDSCVMWNVVDVNSAGERSREFSWSERNVSLKHIMQHIEAAPDIMHYALLGLRKWS SKTRASEVQEPFSRCHVHNFIILNVDLTQNVQYNQNRFLCDDVDFNLRVHSAGLLLCRFNRFSVMKKQ IVVGGHRSFHITSKVSDNSAAVVPAQYICAPDSKHTFLAAPAQLLLEKFLQHHSHLFFPLSLKNHDHP VLSVDCYLNLGSQISVCYVSSRPHSLNISCSDLLFSGLLLYLCDSFVGASFLKKFHFLKGATLCVICQ DRSSLRQTVVRLELEDEWQFRLRDEFQTANAREDRPLFFLTGRHI		

Further analysis of the NOV1a protein yielded the following properties shown in Table 1B.

Table 1B. Protein Sequence Properties NOV1a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1C.

Table 1C. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG61876	Prostate cancer-associated protein #77 - Mammalia, 947 aa. [WO200230268-A2, 18-APR-2002]	1003..1949 1..947	946/947 (99%) 947/947 (99%)	0.0
AAB95517	Human protein sequence SEQ ID NO:18089 - Homo sapiens, 875 aa. [EP1074617-A2, 07-FEB-2001]	775..1606 59..854	399/835 (47%) 534/835 (63%)	0.0
AAO04442	Human polypeptide SEQ ID NO 18334 - Homo sapiens, 112 aa. [WO200164835-A2, 07-SEP-2001]	1190..1301 1..112	110/112 (98%) 110/112 (98%)	5e-56
ABG00933	Novel human diagnostic protein #924 - Homo sapiens, 172 aa. [WO200175067-A2, 11-OCT-2001]	109..258 2..145	101/150 (67%) 115/150 (76%)	9e-51
ABG07439	Novel human diagnostic protein #7430 - Homo sapiens, 175 aa. [WO200175067-A2, 11-OCT-2001]	1223..1348 4..131	61/128 (47%) 75/128 (57%)	5e-24

In a BLAST search of public sequence datbases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1D.

Table 1D. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JLG7	Kiaa0575 - Mus musculus (Mouse), 1954 aa.	1..1949 1..1954	1729/1957 (88%) 1818/1957 (92%)	0.0
Q9H2Q8	GREB1a - Homo sapiens (Human), 1001 aa (fragment).	1..1001 1..1001	999/1001 (99%) 999/1001 (99%)	0.0
O60321	KIAA0575 protein - Homo sapiens (Human), 947 aa.	1003..1949 1..947	946/947 (99%) 947/947 (99%)	0.0

Q9CYA3	8 days embryo cDNA, RIKEN full-length enriched library, clone:5730583K22, full insert sequence - Mus musculus (Mouse), 511 aa.	1439..1949 1..511	471/511 (92%) 492/511 (96%)	0.0
Q9H2Q7	GREB1b - Homo sapiens (Human), 457 aa.	1..449 1..449	448/449 (99%) 448/449 (99%)	0.0

PFam analysis predicts that the NOV1a protein contains the domains shown in the Table 1E.

Table 1E. Domain Analysis of NOV1a

Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C4	1898..1908	5/11 (45%) 10/11 (91%)	0.6

Example 2.

The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 2A.

Table 2A. NOV2 Sequence Analysis

	SEQ ID NO: 3	4995 bp	
NOV2a, CG106287-01 DNA Sequence	GCCGCGCCGAGGAGGCTGCCGCTCTGGCTTGCCGCCCCCCGCCGCCGCTGCACACCGGACCCAGCCGC CGTGCCGCGGGCC**ATG**GACCTGCCCAGGGGCCTGGTGGTGGCCTGGGCGCTCAGCCTGTGGCCAGGGT TCACGGACACCTTCAACATGGACACCAGGAAGCCCCGGGTCATCCCTGGCTCCAGGACCGCCTTCTTT GGCTACACAGTGCAGCAGCACGACATCAGTGGCAATAAGTGGCTGGTCGTGGGCGCCCCACTGGAAAC CAATGGCTACCAGAAGACGGGAGACGTGTACAAGTGTCCAGTGATCCACGGGAACTGCACCAAACTCA ACCTGGGAAGGGTCACCCTGTCCAACGTGTCCGAGCGGAAAGACAACATGCGCCTCGGCCTTAGTCTC GCCACCAACCCCAAGGACAACAGCTTCCTGGCCTGCAGCCCCCTCTGGTCTCATGAGTGTGGGAGCTC CTACTACACCACAGGGATGTGTTCAAGAGTCAACTCCAACTTCAGGTTCTCCAAGACCGTGGCCCCAG CTCTCCAAAGGTGCCAGACCTACATGGACATCGTCATTGTCCTGGATGGCTCCAACAGCATCTACCCC TGGGTGGAGGTTCAGCACTTCCTCATCAACATCCTGAAAAAGTTTTACATTGGCCCAGGGCAGATCCA GGTTGGAGTTGTGCAGTATGGCGAAGATGTGGTGCATGAGTTTCACCTCAATGACTACAGGTCTGTAA AAGATGTGGTGGAAGCTGCCAGCCACATTGAGCAGAGAGGAGGAACAGAGACCCGGACGGCATTTGGC ATTGAATTTGCACGCTCAGAGGCTTTCCAGAAGGGTGGAAGGAAAGGAGCCAAGAAGGTGATGATTGT CATCACAGATGGGGAGTCCCACGACAGCCCAGACCTGGAGAAGGTGATCCAGCAAAGCGAAAGAGACA ACGTAACAAGATATGCGGTGGCCGTCCTGGGCTACTACAACCGCAGGGGGATCAATCCAGAAACTTTT CTAAATGAAATCAAATACATCGCCAGTGACCCTGATGACAAGCACTTCTTCAATGTCACTGATGAGGC TGCCTTGAAGGACATTGTCGATGCCCTGGGGGACAGAATCTTCAGCCTGGAAGGCACCAACAAGAACG AGACCTCCTTTGGGCTGGAGATGTCACAGACGGGCTTTTCCTCGCACGTGGTGGAGGATGGGGTTCTG CTGGGAGCCGTCGGTGCCTATGACTGGAATGGAGCTGTGCTAAAGGAGACGAGTGCCGGGAAGGTCAT TCCTCTCCGCGAGTCCTACCTGAAAGAGTTCCCCGAGGAGCTCAAGAACCATGGTGCATACCTGGGGT ACACAGTCACATCGGTCGTGTCCTCCAGGCAGGGGCGAGTGTACGTGGCCGGAGCCCCCCGGTTCAAC CACACGGGCAAGGTCATCCTGTTCACCATGCACAACAACCGGAGCCTCACCATCCACCAGGCTATGCG GGGCCAGCAGATAGGCTCTTACTTTGGGAGTGAAATCACCTCGGTGGACATCGACGGCGACGGCGTGA CTGATGTCCTGCTGGTGGGCGCACCCATGTACTTCAACGAGGGCCGTGAGCGAGGCAAGGTGTACGTC TATGAGCTGAGACAGAACCGGTTTGTTTATAACGGAACGCTAAAGGATTCACACAGTTACCAGAATGC CCGATTTGGGTCCTCCATTGCCTCAGTTCGAGACCTCAACCAGGATTCCTACAATGACGTGGTGGTGG GAGCCCCCCTGGAGGACAACCACGCAGGAGCCATCTACATCTTCCACGGCTTCCGAGGCAGCATCCTG AAGACACCTAAGCAGAGAATCACAGCCTCAGAGCTGGCTACCGGCCTCCAGTATTTTGGCTGCAGCAT		

	CCACGGGCAATTGGACCTCAATGAGGATGGGCTCATCGACCTGGCAGTGGGAGCCCTTGGCAACGCTG TGATTCTGTGGTCCCGCCCAGTGGTTCAGATCAATGCCAGCCTCCACTTTGAGCCATCCAAGATCAAC ATCTTCCACAGAGACTGCAAGCGCAGTGGCAGGGATGCCACCTGCCTGGCCGCCTTCCTCTGCTTCAC GCCCATCTTCCTGGCACCCCATTTCCAAACAACAACTGTTGGCATCAGATACAACGCCACCATGGATG AGAGGCGGTATACACCGAGGGCCCACCTGGACGAGGGCGGGGACCGATTCACCAACAGAGCCGTACTG CTCTCCTCCGGCCAGGAGCTCTGTGAGCGGATCAACTTCCATGTCCTGGACACTGCTGACTACGTGAA GCCAGTGACCTTCTCAGTCGAGTATTCCCTGGAGGACCCTGACCATGGCCCCATGCTGGACGACGGCT GGCCCACCACTCTCAGAGTCTCGGTGCCCTTCTGGAACGGCTGCAATGAGGATGAGCACTGTGTCCCT GACCTTGTGTTGGATGCCCGGAGTGACCTGCCCACGGCCATGGAGTACTGCCAGAGGGTGCTGAGGAA GCCTGCGCAGGACTGCTCCGCATACACGCTGTCCTTCGACACCACAGTCTTCATCATAGAGAGCACAC GCCAGCGAGTGGCGGTGGAGGCCACACTGGAGAACAGGGGCGAGAACGCCTACAGCACGGTCCTAAAT ATCTCGCAGTCAGCAAACCTGCAGTTTGCCAGCTTGATCCAGAAGGAGGACTCAGACGGTAGCATTGA GTGTGTGAACGAGGAGAGGAGGCTCCAGAAGCAAGTCTGCAACGTCAGCTATCCCTTCTTCCGGGCCA AGGCCAAGGTGGCTTTCCGTCTTGATTTTGAGTTCAGCAAATCCATCTTCCTACACCACCTGGAGATC GAGCTCGCTGCAGGCAGTGACAGTAATGAGCGGGACAGCACCAAGGAAGACAACGTGGCCCCCTTACG CTTCCACCTCAAATACGAGGCTGACGTCCTCTTCACCAGGAGCAGCAGCCTGAGCCACTACGAGGTCA AGCTCAACAGCTCGCTGGAGAGATACGATGGTATCGGGCCTCCCTTCAGCTGCATCTTCAGGATCCAG AACTTGGGCTTGTTCCCCATCCACGGGATTATGATGAAGATCACCATTCCCATCGCCACCAGGAGCGG CAACCGCCTACTGAAGCTGAGGGACTTCCTCACGGACGAGGTAGCGAACACGTCCTGTAACATCTGGG GCAATAGCACTGAGTACCGGCCCACCCCAGTGGAGGAAGACTTGCGTCGTGCTCCACAGCTGAATCAC AGCAACTCTGATGTCGTCTCCATCAACTGCAATATACGGCTGGTCCCCAACCAGGAAATCAATTTCCA TCTACTGGGGAACCTGTGGTTGAGGTCCCTAAAAGCACTCAAGTACAAATCCATGAAAATCATGGTCA ACGCAGCCTTGCAGAGGCAGTTCCACAGCCCCTTCATCTTCCGTGAGGAGGATCCCAGCCGCCAGATC GTGTTTGAGATCTCCAAGCAAGAGGACTGGCAGGTCCCCATCTGGATCATTGTAGGCAGCACCCTGGG GGGCCTCCTACTGCTGGCCCTGCTGGTCCTGGCACTGTGGAAGCTCGGCTTCTTTAGAAGTGCCAGGC GCAGGAGGGAGCCTGGTCTGGACCCCACCCCCAAAGTGCTGGAG**TGA**GGCTCCAGAGGAGACTTTGAG TTGATGGGGGCCAGGACACCAGTCCAGGTAGTGTTGAGACCCAGGCCTGTGGCCCCACCGAGCTGGAG CGGAGAGGAAGCCAGCTGGCTTTGCACTTGACCTCATCTCCCGAGCAATGGCGCCTGCTCCCTCCAGA ATGGAACTCAAGCTGGTTTTAAGTGGAACTGCCCTACTGGGAGACTGGGACACCTTTAACACAGACCC CTAGGGATTTAAAGGACACCCCTACACACACCCAGGCCCACGCCAAGGCCTCCCTCAGGCTCTGTGG AGGGCATTTGCTGCCCCAGCTACTAAGGTGCTAGGAATTCGTAATCATCCCCATCCTCCAGAGAAACC CAGGGAGGAAGACTGTAAATACGAACCCAATCTGCACACTCCAGGCCTCTAGTTCCAGAAGGATCCAA GACAAAACAGATCTGAATTCTGCCCTTTTCTCTCACCCATCCCACCCCTCCATTGGCTCCCAAGTCAC ACCCACTCCCTTCCCCATAGATAGGCCCCTGGGGCTCCCGAAGAATGAACCCAAGAGCAAGGGCTTGA TGGTGACAGCTGCAAGCCAGGGATGAAGAAAGACTCTGAGATGTGGAGACTGATGGCCAGGCAAGTGG GACCAGGATACTGGACGCTGTCCTGAGATGAGAGGTAGCCGGGCTCTGCACCCACGTGCATTCACATT GACCGCAACTCACACATTCCCCCACCAGCTGCAGCCCCTTGCTCTCAGCTGCCAACCCTCCCGGGTCA CTTTTGTTCCCAGGTACCTCATGGGAAGCATGTGGATGACACAATCCCTGGGGCTGTGCATTCCCACG TCTTCTTGCTGCAGCCTGCCCCTAGACATGGACGCACCGGCCTGGCTGCAGCTGGGCAGCAGGGGTAG GGGTAGGGAGCCTCCCCTCCCTGTATCACCCCCTCCCTACACACACACACACACACACACACACACAC TGCCTCCCATCCTTCCCTCATGCCCGCCAGTGCACAGGGAAGGGCTTGGCCAGCGCTGTTGAGGGGTC CCCTCTGGAATGCACTGAATAAAGCACGTGCAAGGACTCCCGGAGCCTGTGCAGCCTTGGTGGCAAAT ATCTCATCTGCCGGCCCCCAGGACAAGTGGTATGACCAGTGATAATGCCCCAAGGACAAGGGGCGTGC CTGGCGCCCAGTGGAGTAATTTATGCCTTAGTCTTGTTTTGAGGTAGAAATGCAAGGGGGACACATGA AAGGCATCAGTCCCCCTGTGCATAGTACGACCTTTACTGTCGTATTTTTGAAAAATTAAAAATACAGT GTTTAAAAACAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 82		ORF Stop: TGA at 3649

	SEQ ID NO: 4	1189 aa	MW at 133608.9kD
NOV2a, CG106287-01 Protein Sequence	MDLPRGLVVAWALSLWPGFTDTFNMDTRKPRVIPGSRTAFFGYTVQQHDISGNKWLVVGAPLETNGYQ KTGDVYKCPVIHGNCTKLNLGRVTLSNVSERKDNMRLGLSLATNPKDNSFLACSPLWSHECGSSYYTT GMCSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGSNSIYPWVEVQHFLINILKKFYIGPGQIQVGVV QYGEDVVHEFHLNDYRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGGRKGAKKVMIVITDG ESHDSPDLEKVIQQSERDNVTRYAVAVLGYYNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAALKD IVDALGDRIFSLEGTNKNETSFGLEMSQTGFSSHVVEDGVLLGAVGAYDWNGAVLKETSAGKVIPLRE SYLKEFPEELKNHGAYLGYTVTSVVSSRQGRVYVAGAPRFNHTGKVILFTMHNNRSLTIHQAMRGQQI GSYFGSEITSVDIDGDGVTDVLLVGAPMYFNEGRERGKVYVYELRQNRFVYNGTLKDSHSYQNARFGS SIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQRITASELATGLQYFGCSIHGQL DLNEDGLIDLAVGALGNAVILWSRPVVQINASLHFEPSKINIFHRDCKRSGRDATCLAAFLCFTPIFL APHFQTTTVGIRYNATMDERRYTPRAHLDEGGDRFTNRAVLLSSGQELCERINFHVLDTADYVKPVTF SVEYSLEDPDHGPMLDDGWPTTLRVSVPFWNGCNEDEHCVPDLVLDARSDLPTAMEYCQRVLRKPAQD CSAYTLSFDTTVFIIESTRQRVAVEATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIECVNE ERRLQKQVCNVSYPFFRAKAKVAFRLDFEFSKSIFLHHLEIELAAGSDSNERDSTKEDNVAPLRFHLK YEADVLFTRSSSLSHYEVKLNSSLERYDGIGPPFSCIFRIQNLGLFPIHGIMMKITIPIATRSGNRLL KLRDFLTDEVANTSCNIWGNSTEYRPTPVEEDLRRAPQLNHSNSDVVSINCNIRLVPNQEINFHLLGN LWLRSLKALKYKSMKIMVNAALQRQFHSPFIFREEDPSRQIVFEISKQEDWQVPIWIIVGSTLGGLLL LALLVLALWKLGFFRSARRRREPGLDPTPKVLE		

	SEQ ID NO: 5	4779 bp	

NOV2b, CG106287-02 DNA Sequence	AGGAGGCTGCCGCTCTGGCTTGCCGCCCCCCGCCGCCGCTGCACACCGGACCCAGCCGCCGTGCCGC GGGCCATGGACCTGCCCAGGGGCCTGGTGGTGGCCTGGGCGCTCAGCCTGTGGCCAGGTTTCACGGA CACCTTCAACATGGACACCAGGAAGCCCCGGGTCATCCCTGGCTCCAGGACCGCCTTCTTTGGCTAC ACAGTGCAGCAGCACGACATCAGTGGCAATAAGTGGCTGGTCGTGGGCGCCCCACTGGAAACCAATG GCTACCAGAAGACGGGAGACGTGTACAAGTGTCCAGTGATCCACGGGAACTGCACCAAACTCAACCT GGGGTGCCAGACCTACATGGACATCGTCATTGTCCTGGATGGCTCCAACAGCATCTACCCCTGGGTG GAGGTTCAGCACTTCCTCATCAACATCCTGAAAAAGTTTTACATTGGCCCAGGGCAGATCCAGGTTG GAGTTGTGCAGTATGGCGAAGATGTGGTGCATGAGTTTCACCTCAATGACTACAGGTCTGTAAAAGA TGTGGTGGAAGCTGCCAGCCACATTGAGCAGAGAGGAGGAACAGAGACCCGGACGGCATTTGGCATT GAATTTGCACGCTCAGAGGCTTTCCAGAAGGGTGGAAGGAAAGGAGCCAAGAAGGTGATGATTGTCA TCACAGATGGGGAGTCCCACGACAGCCCAGACCTGGAGAAGGTGATCCAGCAAAGCGAAAGAGACAA CGTAACAAGATATGCGGTGGCCGTCCTGGGCTACTACAACCGCAGGGGGATCAATCCAGAAACTTTT CTAAATGAAATCAAATACATCGCCAGTGACCCTGATGACAAGCACTTCTTCAATGTCACTGATGAGG CTGCCTTGAAGGACATTGTCGATGCCCTGGGGGACAGAATCTTCAGCCTGGAAGGCACCAACAAGAA CGAGACCTCCTTTGGGCTGGAGATGTCACAGACGGGCTTTTCCTCGCACGTGGTGGAGGATGGGGTT CTGCTGGGAGCCGTCGGTGCCTATGACTGGAATGGAGCTGTGCTAAAGGAGACGAGTGCCGGGAAGG TCATTCCTCTCCGCGAGTCCTACCTGAAAGAGTTCCCCGAGGAGCTCAAGAACCATGGTGCATACCT GGGGTACACAGTCACATCGGTCGTGTCCTCCAGGCAGGGGCGAGTGTACGTGGCCGGAGCCCCCCGG TTCAACCACACGGGCAAGGTCATCCTGTTCACCATGCACAACAACCGGAGCCTCACCATCCACCAGG CTATGCGGGGCCAGCAGATAGGCTCTTACTTTGGGAGTGAAATCACCTCGGTGGACATCGACGGCGA CGGCGTGACTGATGTCCTGCTGGTGGGCGCACCCATGTACTTCAACGAGGGCCGTGAGCGAGGCAAG GTGTACGTCTATGAGCTGAGACAGAACCGGTTTGTTTATAACGGAACGCTAAAGGATTCACACAGTT ACCAGAATGCCCGATTTGGGTCCTCCATTGCCTCAGTTCGAGACCTCAACCAGGATTCCTACAATGA CGTGGTGGTGGGAGCCCCCCTGGAGGACAACCACGCAGGAGCCATCTACATCTTCCACGGCTTCCGA GGCAGCATCCTGAAGACACCTAAGCAGAGAATCACAGCCTCAGAGCTGGCTACCGGCCTCCAGTATT TTGGCTGCAGCATCCACGGGCAATTGGACCTCAATGAGGATGGGCTCATCGACCTGGCAGTGGGAGC CCTTGGCAACGCTGTGATTCTGTGGTCCCGCCCAGTGGTTCAGATCAATGCCAGCCTCCACTTTGAG CCATCCAAGATCAACATCTTCCACAGAGACTGCAAGCGCAGTGGCAGGGATGCCACCTGCCTGGCCG CCTTCCTCTGCTTCACGCCCATCTTCCTGGCACCCCATTTCCAAACAACAACTGTTGGCATCAGATA CAACGCCACCATGGATGAGAGGCGGTATACACCGAGGGCCCACCTGGACGAGGGCGGGGACCGATTC ACCAACAGAGCCGTACTGCTCTCCTCCGGCCAGGAGCTCTGTGAGCGGATCAACTTCCATGTCCTGG ACACTGCTGACTACGTGAAGCCAGTGACCTTCTCAGTCGAGTATTCCCTGGAGGACCCTGACCATGG CCCCATGCTGGACGACGGCTGGCCCACCACTCTCAGAGTCTCGGTGCCCTTCTGGAACGGCTGCAAT GAGGATGAGCACTGTGTCCCTGACCTTGTGTTGGATGCCCGGAGTGACCTGCCCACGGCCATGGAGT ACTGCCAGAGGGTGCTGAGGAAGCCTGCGCAGGACTGCTCCGCATACACGCTGTCCTTCGACACCAC AGTCTTCATCATAGAGAGCACACGCCAGCGAGTGGCGGTGGAGGCCACACTGGAGAACAGGGGCGAG AACGCCTACAGCACGGTCCTAAATATCTCGCAGTCAGCAAACCTGCAGTTTGCCAGCTTGATCCAGA AGGAGGACTCAGACGGTAGCATTGAGTGTGTGAACGAGGAGAGGAGGCTCCAGAAGCAAGTCTGCAA CGTCAGCTATCCCTTCTTCCGGGCCAAGGCCAAGGTGGCTTTCCGTCTTGATTTTGAGTTCAGCAAA TCCATCTTCCTACACCACCTGGAGATCGAGCTCGCTGCAGGCAGTGACAGTAATGAGCGGGACAGCA CCAAGGAAGACAACGTGGCCCCCTTACGCTTCCACCTCAAATACGAGGCTGACGTCCTCTTCACCAG GAGCAGCAGCCTGAGCCACTACGAGGTCAAGCTCAACAGCTCGCTGGAGAGATACGATGGTATCGGG CCTCCCTTCAGCTGCATCTTCAGGATCCAGAACTTGGGCTTGTTCCCCATCCACGGGATTATGATGA AGATCACCATTCCCATCGCCACCAGGAGCGGCAACCGCCTACTGAAGCTGAGGGACTTCCTCACGGA CGAGGTAGCGAACACGTCCTGTAACATCTGGGGCAATAGCACTGAGTACCGGCCCACCCCAGTGGAG GAAGACTTGCGTCGTGCTCCACAGCTGAATCACAGCAACTCTGATGTCGTCTCCATCAACTGCAATA TACGGCTGGTCCCCAACCAGGAAATCAATTTCCATCTACTGGGGAACCTGTGGTTGAGGTCCCTAAA AGCACTCAAGTACAAATCCATGAAAATCATGGTCAACGCAGCCTTGCAGAGGCAGTTCCACAGCCCC TTCATCTTCCGTGAGGAGGATCCCAGCCGCCAGATCGTGTTTGAGATCTCCAAGCAAGAGGACTGGC AGGTCCCCATCTGGATCATTGTAGGCAGCACCCTGGGGGGCCTCCTACTGCTGGCCCTGCTGGTCCT GGCACTGTGGAAGCTCGGCTTCTTTAGAAGTGCCAGGCGCAGGAGGGAGCCTGGTCTGGACCCCACC CCCAAAGTGCTGGAGTGAGGCTCCAGAGGAGACTTTGAGTTGATGGGGGCCAGGACACCAGTCCAGG TAGTGTTGAGACCCAGGCCTGTGGCCCCACCGAGCTGGAGCGGAGAGGAAGCCAGCTGGCTTTGCAC TTGACCTCATCTCCCGAGCAATGGCGCCTGCTCCCTCCAGAATGGAACTCAAGCTGGTTTTAAGTGG AACTGCCCTACTGGGAGACTGGGACACCTTTAACACAGACCCCTAGGGATTTAAAGGACACCCCTA CACACACCCAGGCCCACGCCAAGGCCTCCCTCAGGCTCTGTGGAGGGCATTTGCTGCCCCAGCTACT AAGGTGCTAGGAATTCGTAATCATCCCCATCCTCCAGAGAAACCCAGGGAGGAAGACTGTAAATACG AACCCAATCTGCACACTCCAGGCCTCTAGTTCCAGAAGGATCCAAGACAAAACAGATCTGAATTCTG CCCTTTTCTCTCACCCATCCCACCCCTCCATTGGCTCCCAAGTCACACCCACTCCCTTCCCCATAGA TAGGCCCCTGGGGCTCCCGAAGAATGAACCCAAGAGCAAGGGCTTGATGGTGACAGCTGCAAGCCAG GGATGAAGAAAGACTCTGAGATGTGGAGACTGATGGCCAGGCAAGTGGGACCAGGATACTGGACGCT GTCCTGAGATGAGAGGTAGCCGGGCTCTGCACCCACGTGCATTCACATTGACCGCAACTCACACATT CCCCCACCAGCTGCAGCCCCTTGCTCTCAGCTGCCAACCCTCCCGGGTCACTTTTGTTCCCAGGTAC CTCATGGGAAGCATGTGGATGACACAATCCCTGGGGCTGTGCATTCCCACGTCTTCTTGCTGCAGCC TGCCCCTAGACATGGACGCACCGGCCTGGCTGCAGCTGGGCAGCAGGGGTAGGGGTAGGGAGCCTCC CCTCCCTGTATCACCCCCTCCCTACACACACACACACACACACACACACACTGCCTCCCATCCTT CCCTCATGCCCGCCAGTGCACAGGGAAGGGCTTGGCCAGCGCTGTTGAGGGGTCCCCTCTGGAATGC ACTGAATAAAGCACGTGCAAGGACTCCCGGAGCCTGTGCAGCCTTGGTGGCAAATATCTCATCTGCC GGCCCCCAGGACAAGTGGTATGACCAGTGATAATGCCCCAAGGACAAGGGGCGTGCCTGGCGCCCAG TGGAGTAATTTATGCCTTAGTCTTGTTTTGAGGTAGAAATGCAAGGGGGACACATGAAAGGCATCAG TCCCCCTGTGCATAGTACGACCTTTACTGTCGTATTTTTGAAAAATTAAAAATACAGTGTTTAAAAA CAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 73		ORF Stop: TGA at 3433

	SEQ ID NO: 6	1120 aa	MW at 125924.3kD
NOV2b, CG106287-02 Protein Sequence	MDLPRGLVVAWALSLWPGFTDTFNMDTRKPRVIPGSRTAFFGYTVQQHDISGNKWLVVGAPLETNGY QKTGDVYKCPVIHGNCTKLNLGCQTYMDIVIVLDGSNSIYPWVEVQHFLINILKKFYIGPGQIQVGV VQYGEDVVHEFHLNDYRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGGRKGAKKVMIVIT DGESHDSPDLEKVIQQSERDNVTRYAVAVLGYYNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAA LKDIVDALGDRIFSLEGTNKNETSFGLEMSQTGFSSHVVEDGVLLGAVGAYDWNGAVLKETSAGKVI PLRESYLKEFPEELKNHGAYLGYTVTSVVSSRQGRVYVAGAPRFNHTGKVILFTMHNNRSLTIHQAM RGQQIGSYFGSEITSVDIDGDGVTDVLLVGAPMYFNEGRERGKVYVYELRQNRFVYNGTLKDSHSYQ NARFGSSIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQRITASELATGLQYFG CSIHGQLDLNEDGLIDLAVGALGNAVILWSRPVVQINASLHFEPSKINIFHRDCKRSGRDATCLAAF LCFTPIFLAPHFQTTTVGIRYNATMDERRYTPRAHLDEGGDRFTNRAVLLSSGQELCERINFHVLDT ADYVKPVTFSVEYSLEDPDHGPMLDDGWPTTLRVSVPFWNGCNEDEHCVPDLVLDARSDLPTAMEYC QRVLRKPAQDCSAYTLSFDTTVFIIESTRQRVAVEATLENRGENAYSTVLNISQSANLQFASLIQKE DSDGSIECVNEERRLQKQVCNVSYPFFRAKAKVAFRLDFEFSKSIFLHHLEIELAAGSDSNERDSTK EDNVAPLRFHLKYEADVLFTRSSSLSHYEVKLNSSLERYDGIGPPFSCIFRIQNLGLFPIHGIMMKI TIPIATRSGNRLLKLRDFLTDEVANTSCNIWGNSTEYRPTPVEEDLRRAPQLNHSNSDVVSINCNIR LVPNQEINFHLLGNLWLRSLKALKYKSMKIMVNAALQRQFHSPFIFREEDPSRQIVFEISKQEDWQV PIWIIVGSTLGGLLLLALLVLALWKLGFFRSARRRREPGLDPTPKVLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 2B.

Table 2B. Comparison of NOV2a against NOV2b.		
Protein Sequence	**NOV2a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV2b	159..1189 90..1120	1017/1031 (98%) 1017/1031 (98%)

Further analysis of the NOV2a protein yielded the following properties shown in Table 2C.

Table 2C. Protein Sequence Properties NOV2a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 23 and 24

A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 2D.

Table 2D. Geneseq Results for NOV2a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV2a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**

AAB25582	ITGA11 protein encoded by human secreted protein gene #7 - Homo sapiens, 1189 aa. [WO200029435-A1, 25-MAY-2000]	1..1189 1..1189	1189/1189 (100%) 1189/1189 (100%)	0.0
ABG12949	Novel human diagnostic protein #12940 - Homo sapiens, 1189 aa. [WO200175067-A2, 11-OCT-2001]	1..1189 1..1189	1188/1189 (99%) 1189/1189 (99%)	0.0
AAU10551	Human A259 polypeptide - Homo sapiens, 1188 aa. [WO200181414-A2, 01-NOV-2001]	1..1189 1..1188	1186/1189 (99%) 1187/1189 (99%)	0.0
AAB50085	Human A259 - Homo sapiens, 1188 aa. [WO200073339-A1, 07-DEC-2000]	1..1189 1..1188	1186/1189 (99%) 1187/1189 (99%)	0.0
AAU14231	Human novel protein #102 - Homo sapiens, 1188 aa. [WO200155437-A2, 02-AUG-2001]	1..1189 1..1188	1186/1189 (99%) 1187/1189 (99%)	0.0

In a BLAST search of public sequence datbases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2E.

Table 2E. Public BLASTP Results for NOV2a

Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UKX5	Integrin alpha-11 precursor - Homo sapiens (Human), 1189 aa.	1..1189 1..1189	1189/1189 (100%) 1189/1189 (100%)	0.0
CAD28200	Sequence 1 from Patent WO0181414 - Homo sapiens (Human), 1188 aa.	1..1189 1..1188	1186/1189 (99%) 1187/1189 (99%)	0.0
CAD28203	Sequence 19 from Patent WO0181414 - Mus musculus (Mouse), 1188 aa.	1..1189 1..1188	1073/1189 (90%) 1130/1189 (94%)	0.0
Q8WYI8	MSTP018 - Homo sapiens (Human), 823 aa.	366..1189 1..823	822/824 (99%) 823/824 (99%)	0.0
O75578	Integrin alpha-10 precursor - Homo sapiens (Human), 1167 aa.	1..1170 1..1150	513/1181 (43%) 723/1181 (60%)	0.0

PFam analysis predicts that the NOV2a protein contains the domains shown in the Table 2F.

Table 2F. Domain Analysis of NOV2a

Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
FG-GAP	38..94	19/65 (29%) 39/65 (60%)	2e-08
vwa	164..345	65/208 (31%) 155/208 (75%)	8.1e-54
FG-GAP	422..475	13/65 (20%) 42/65 (65%)	4.2e-06
FG-GAP	477..537	23/65 (35%) 48/65 (74%)	2.6e-12
FG-GAP	539..598	24/67 (36%) 53/67 (79%)	1.6e-15
FG-GAP	601..653	20/66 (30%) 42/66 (64%)	3.2e-09

Example 3.

The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis

	SEQ ID NO: 7	1915 bp	
NOV3a, CG106417-01 DNA Sequence	CCCGGGGACCCGCCGCCGCCGGTCATGTGGCCGGACTGCTCCTTCGGCCGCCTGTGTCGCGCTCC TGCTGCCGGGGGCACCAGCCCGAGGCTACACCGGAGGAAGCCGCCCGGGCACTTCGCGGCCGAGAGG AGACGCCGACTGGGCCCCCACGTCTGCCTCTCTGGGTTTGGGAGTGGCTGCTGCCCTGGCTGGGCGCC CTCTATGGGTGGTGGGCACTGCACCCTGCGTAAGCTCTGCTCCTTCGGCTGTGGGAGTGGCATCTGCA TCGCTCCCAATGTCTGCTCCTGCCAGGATGGAGAGCAAGGGGCCACCTGCCCAGAAACCCATGGACCA TGTGGGGAGTACGGCTGTGACCTTACCTGCAACCATGGAGGCTGTCAGGAGGTGGCCCGAGTGTGCCC CGTGGGCTTCTCGATGACGGAGACAGCTGTTGGCATCAGGTGTACAGACATTGACGAATGTGTAACCT CCTCCTGCGAGGGCCACTGTGTGAACACAGAAGGTGGGTTTGTGTGCGAGTGTGGGCCGGGCATGCAG CTGTCTGCCGACCGCCACAGCTGCCAAGACACTGACGAATGCCTAGGGACTCCCTGTCAGCAGAGATG TAAAAACAGCATTGGCAGCTACAAGTGTTCCTGTCGAACTGGCTTCCACCTTCATGGCAACCGGCACT CCTGTGTAGACGCAAACGAGTGTCGGACGCCATCGGAGACGCGAGTCTGTCACCATTCCTGCCACAAC ACCGTGGGCAGCTTCGTATGCACATGCGGACCTGGTTTCAGGTTCGGAGCTGACCGCGTGTCTGTTTC AGCTTTCCCGAAAGCCGTGCTGGCCCCATCTGCCATCCTGCAACCCCGGCAACACCCGTCCAAGATGC TTCTGTTGCTTCCTGAGGCCGGCCGGCCTGCCCTGTCCCCAGGACATAGCCCTCCTTCTGGGGCTCCA GGGCCCCCAGCCGGAGTCAGGACCACCCGCCTGCCATCTCCCACCCCACGACTACCCACATCCTCCCC TTCTGCCCTGCTGGCCACCCCAGTGCCTACTGCCTCCCTGCTGGGGAACCTCAGACCCCCCTCACTCC TTCAGGGGGAGGTGATGGGGACCCCTTCCTCACCCAGGGGCCCTGAGTCCCCCCGACTGGCAGCAGGG CCCTCTCCCTGCTGGCACCTGGGAGCCATGCATGAATCAAGGAGTCGCTGGACAGAGCCTGGGTGTTC CCAGTGCTGGTGCGAGGTGGGTGGGCCTTGTGGTGGCGACGGGAAGGTGACCTGTGAAAAGGTGAGGT GTGAAGCTGCTTGTTCCCACCCAATTCCCTCCAGAGATGGTGGGTGCTGCCCATCGTGCACAGGTTCC TATTTGTCCTTCAAAGGCTGTTTTCACAGTGGTGTCGTCCGAGCTGAAGGGGATGTGTTTTCACCTCC CAATGAGAACTGCACCGTCTGTGTCTGTCTGGCTGGAAACGTGTCGTGCATGTTTCGTGAGTGTCCTT TTGGCCCGTGTGAGACCCCCCATAAAGACAGATGCTATTTCCACGGCCGGTGGTACGCAGACGGGGCT		

	GTGTTCAGTGGGGGTGGTGACGAGTGTACCACCTGTGTTTGCCAGCAGAATGGGGAGGTGGAGTGCTC CTTCATGCCCTGCCCTGAGCTGGCCTGCCCCCGAGAAGAGTGGCGGCTGGGCCCTGGGCAGTGTTGCT TCACCTGCCAGGAGCCCACACCCTCGACAGGTGGCTGCTCTCTTGACGACAACGGGGTTGAGTTTCCG ATTGGACAGATCTGGTCGCCTGGTGACCCCTGTCCAGGCAGATGGCTCGGTGAGCTGCAAGAGGACAG ACTGTGTGGACTCCTGCCCTCACCCGATCCGGATCCCTGGACAGTGCTGCCCAGACTGTTCAGCAGGT AATCCCCTGCC		
	ORF Start: ATG at 26		ORF Stop: TAA at 1904

	SEQ ID NO: 8	626 aa	MW at 66006.2kD
NOV3a, CG106417-01 Protein Sequence	MWAGLLLRAACVALLLPGAPARGYTGRKPPGHFAAERRRRLGPHVCLSGFGSGCCPGWAPSMGGGHCT LRKLCSFGCGSGICIAPNVCSCQDGEQGATCPETHGPCGEYGCDLTCNHGGCQEVARVCPVGFSMTET AVGIRCTDIDECVTSSCEGHCVNTEGGFVCECGPGMQLSADRHSCQDTDECLGTPCQQRCKNSIGSYK CSCRTGFHLHGNRHSCVDANECRTPSETRVCHHSCHNTVGSFVCTCGPGFRFGADRVSVSAFPKAVLA PSAILQPRQHPSKMLLLLPEAGRPALSPGHSPPSGAPGPPAGVRTTRLPSPTPRLPTSSPSALLATPV PTASLLGNLRPPSLLQGEVMGTPSSPRGPESPRLAAGPSPCWHLGAMHESRSRWTEPGCSQCWCEVGG PCGGDGKVTCEKVRCEAACSHPIPSRDGGCCPSCTGSYLSFKGCFHSGVVRAEGDVFSPPNENCTVCV CLAGNVSCMFRECPFGPCETPHKDRCYFHGRWYADGAVFSGGGDECTTCVCQQNGEVECSFMPCPELA CPREEWRLGPGQCCFTCQEPTPSTGGCSLDDNGVEFPIGQIWSPGDPCPGRWLGELQEDRLCGLLPSP DPDPWTVLPRLFSR		

	SEQ ID NO: 9	2040 bp	
NOV3b, CG106417-03 DNA Sequence	ATGTGGGCCGGACTGCTCCTTCGGGCCGCCTGTGTCGCGCTCCTGCTGCCGGGGGCACCAGCCCGAG GCTACACCGGGAGGAAGCCGCCCGGGCACTTCGCGGCCGAGAGGAGACGCCGACTGGGCCCCCACGT CTGCCTCTCTGGGTTTGGGAGTGGCTGCTGCCCTGGCTGGGCGCCCTCTATGGGTGGTGGGCACTGC ACCCTGCTCTGCTCCTTCGGCTGTGGAGTGGCATCTGCATCGCTCCCAATGTCTGCTCCTGCCAGG ATGGAGAGCAAGGGGCCGAAACCCATGGACCATGTGGGGAGTACGGCTGTGACCTTACCTGCAACCA TGGAGGCTGTCAGGAGGTGGCCCGAGTGTGCCCCGTGGGCTTCTCGATGACGGAGACAGCTGTTGGC ATCAGGTGTGACATTGACGAATGTGTAACCTCCTCCTGCGAGGGCCACTGTGTGAACACAGAAGGTG GGTTTGTGTGCGAGTGTGGGCCGGGCATGCAGCTGTCTGCCGACCGCCACAGCTGCCAAGACACTGA CGAATGCCTAGGGACTCCCTGTCAGCAGAGATGTAAAAACAGCATTGGCAGCTACAAGTGTTCCTGT CGAACTGGCTTCCACCTTCATGGCAACCGGCACTCCTGTGTAGACGCAAACGAGTGTCGGACGCCAT CGGAGACGCGAGTCTGTCACCATTCCTGCCACAACACCGTGGGCAGCTTCGTATGCACATGCGGACC TGGTTTCAGGTTCGGAGCTGACCGCGTGCCATGTGAAGGTGAGCGCCAGGCCAGAGACCTCCGTGCT TCTGTTTCAGCTTTCCCGAAAGCCGTGCTGGCCCCATCTGCCATCCTGCAACCCCGGCAACACCCGT CCAAGATGCTTCTGTTGCTTCCTGAGGCCGGCCGGCCTGCCCTGTCCCCAGGACATAGCCCTCCTTC TGGGGCTCCAGGGCCCCCAGCCGGAGTCAGGACCACCCGCCTGCCATCTCCCACCCCACGACTACCC ACATCCTCCCCTTCTGCCCCTGTGTGGCTGCTGTCCACCCTGCTGGCCACCCCAGTGCCTACTGCCT CCCTGCTGGGGAACCTCAGACCCCCCTCACTCCTTCAGGGGGAGGTGATGGGGACCCCTTCCTCACC CAGGGGCCCTGAGTCCCCCCGACTGGCACCAGGGCCCTCTCCCTGCTGGCACCTGGGAGCCATGCAT GAATCAAGGAGTCGCTGGACAGAGCCTGGGTGTTCCCAGTGCTGGTGCGAGGGCTCTAACTCCTGCT TGTGCTTCGACGGGAAGGTGACCTGTGAAAAGGTGAGGTGTGAAGCTGCTTGTTCCCACCCAATTCC CTCCAGAGATGGTGGGTGCTGCCCATCGTGCACAGGTGGCTGTTTTCACAGTGGTGTCGTCCGAGCT GAAGGGGATGTGTTTTCACCTCCCAATGAGAACTGCACCGTCTGTGTCTGTCTGGCTGGAAACGTGT CGTGCATGTTTCGTGAGTGTCCTTTTGGCCCGTGTGAGACCCCCCATAAAGACTGCAGGTGCCCACC TGGAAGATGCTATTTCCACGGCCGGTGGTACGCAGACGGGGCTGTGTTCAGTGGGGGTGGTGACGAG TGTACCACCTGTGTTTGCCAGAATGGGGAGGTGGAGTGCTCCTTCATGCCCTGCCCTGAGCTGGCCT GCCCCCGAGAAGAGTGGCGGCTGGGCCCTGGGCAGTGTTGCTTCACCTGCCAGGAGCCCACACCCTC GACAGGTCTTGACGACAACGGGGTTGAGTTTCCGATTGGACAGATCTGGTCGCCTGGTGACCCCTGT GAGAGATGGCTCGGTGAGCTGCAAGAGGACAGACTGTGTGGACTCCTGCCCTCACCCGATCCGGATC CCTGGACAGTGCTGCCCAGACTGTTCAGCAGGTAATCCCCTGCCTCTGCCCCAAGCCCCCAGGGCAG GGCATCTCAGGCATCGGGCTCCTTAAGCCCTATACAGCCTTCATCTCATGTCGTCCTAACAACCCCA AGGGACAACCCCATTGCACAGATAAGGAAA		
	ORF Start: ATG at 1		ORF Stop: TAA at 1909

	SEQ ID NO: 10	636 aa	MW at 67370.7kD
NOV3b, CG106417-03 Protein Sequence	MWAGLLLRAACVALLLPGAPARGYTGRKPPGHFAAERRRRLGPHVCLSGFGSGCCPGWAPSMGGGHC TLLCSFGCGSGICIAPNVCSCQDGEQGAETHGPCGEYGCDLTCNHGGCQEVARVCPVGFSMTETAVG IRCDIDECVTSSCEGHCVNTEGGFVCECGPGMQLSADRHSCQDTDECLGTPCQQRCKNSIGSYKCSC RTGFHLHGNRHSCVDANECRTPSETRVCHHSCHNTVGSFVCTCGPGFRFGADRVPCEGERQARDLRA SVSAFPKAVLAPSAILQPRQHPSKMLLLLPEAGRPALSPGHSPPSGAPGPPAGVRTTRLPSPTPRLP TSSPSAPVWLLSTLLATPVPTASLLGNLRPPSLLQGEVMGTPSSPRGPESPRLAAGPSPCWHLGAMH ESRSRWTEPGCSQCWCEGSNSCLCFDGKVTCEKVRCEAACSHPIPSRDGGCCPSCTGGCFHSGVVRA EGDVFSPPNENCTVCVCLAGNVSCMFRECPFGPCETPHKDCRCPPGRCYFHGRWYADGAVFSGGGDE CTTCVCQNGEVECSFMPCPELACPREEWRLGPGQCCFTCQEPTPSTGLDDNGVEFPIGQIWSPGDPC ERWLGELQEDRLCGLLPSPDPDPWTVLPRLFSR		

	SEQ ID NO: 11	1821 bp	
NOV3c, CG106417-04 DNA Sequence	ATGTGGGCCGGACTGCTCCTTCGGGCCGCCTGTGTCGTCTGCTCCTTCGGCTGTGGGAGTGGCATCTG CATCGCTCCCAATGTCTGCTCCTGCCAGGATGGAGAGCAAGGGGCCGAAACCCATGGACCATGTGGGG AGTACGGCTGTGACCTTACCTGCAACCATGGAGGCTGTCAGGAGGTGGCCCGAGTGTGCCCCGTGGGC TTCTCGATGACGGAGACAGCTGTTGGCATCAGGTGTGACATTGACGAATGTGTAACCTCCTCCTGCGA GGGCCACTGTGTGAACACAGAAGGTGGGTTTGTGTGCGAGTGTGGGCCGGGCATGCAGCTGTCTGCCG ACCGCCACAGCTGCCAAGACACTGACGAATGCCTAGGGACTCCCTGTCAGCAGAGATGTAAAAACAGC ATTGGCAGCTACAAGTGTTCCTGTCGAACTGGCTTCCACCTTCATGGCAACCGGCACTCCTGTGTAGA TGTAAACGAGTGTCGGAGGCCATTGGAGAGGCGAGTCTGTCACCATTCCTGCCACAACACCGTGGGCA GCTTCCTATGCACATGCCGACCTGGCTTCAGGCTCCGAGCTGACCGCGTGTCCTGTGAAGGTGAGCGC CAGGCTTTCCCGAAAGCCGTGCTGGCCCCATCTGCCATCCTGCAACCCCGGCAACACCCGTCCAAGAT GCTTCTGTTGCTTCCTGAGGCCGGCCGGCCTGCCCTGTCCCCAGGACATAGCCCTCCTTCTGGGGCTC CAGGGCCCCCAGCCGGAGTCAGGACCACCCGCCTGCCATCTCCCACCCCACGACTACCCACATCCTCC CCTTCTGCCCCTGTGTGGCTGCTGTCCACCCTGCTGGCCACCCCAGTGCCTACTGCCTCCCTGCTGGG GAACCTCAGACCCCCCTCACTCCTTCAGGGGGAGGTGATGGGGACCCCTTCCTCACCCAGGGGCCCTG AGTCCCCCCGACTGGCAGCAGGGCCCTCTCCCTGCTGGCACCTGGGAGCCATGCATGAATCAAGGAGT CGCTGGACAGAGCCTGGGTGTTCCCAGTGCTGGTGCGAGGACGGGAAGGTGACCTGTGAAAAGGTGAG GTGTGAAGCTGCTTGTTCCCACCCAATTCCCTCCAGAGATGGTGGGTGCTGCCCATCGTGCACAGGTT GTTTTCACAGTGGTGTCGTCCGAGCTGAAGGGGATGTGTTTTCACCTCCCAATGAGAACTGCACCGTC TGTGTCTGTCTGGCTGGAAACGTGTCCTGCATCTCTCCTGAGTGTCCTTCTGGCCCCTGTCAGACCCC CCCACAGACGGATTGCTGTACTTGTGTTCCAGGTAGATGCTATTTCCACGGCCGGTGGTACGCAGACG GGGCTGTGTTCAGTGGGGGTGGTGACGAGTGTACCACCTGTGTTTGCCAGAATGGGGAGGTGGAGTGC TCCTTCATGCCCTGCCCTGAGCTGGCCTGCCCCCGAGAAGAGTGGCGGCTGGGCCCTGGGCAGTGTTG CTTCACCTGCCAGGAGCCCACACCCTCGACAGGTCTTGACGACAACGGGGTTGAGTTTCCGATTGGAC AGATCTGGTCGCCTGGTGACCCCTGTGAGAGATGGCTCGGTGAGCTGCAAGAGGACAGACTGTGTGGA CTCCTGCCCTCACCCGATCCGGATCCCTGGACAGTGCTGCCCAGACTGTTCAGCAGGTAATCCCCTGC CTCTGCCCCAAGCCCCCAGGGCAGGGCATCTCAGGCATCGGGCTCCTTAAGCCCTATACAGCCTTCAT CTCATGTCGTCCTAACAACCCCAAGGGACAACCCCATTGCACAGATAAGGAAA		
	ORF Start: ATG at 1		ORF Stop: TAA at 1690

	SEQ ID NO: 12	563 aa	MW at 59951.3kD
NOV3c, CG106417-04 Protein Sequence	MWAGLLLRAACVVCSFGCGSGICIAPNVCSCQDGEQGAETHGPCGEYGCDLTCNHGGCQEVARVCPVG FSMTETAVGIRCDIDECVTSSCEGHCVNTEGGFVCECGPGMQLSADRHSCQDTDECLGTPCQQRCKNS IGSYKCSCRTGFHLHGNRHSCVDVNECRRPLERRVCHHSCHNTVGSFLCTCRPGFRLRADRVSCEGER QAFPKAVLAPSAILQPRQHPSKMLLLLPEAGRPALSPGHSPPSGAPGPPAGVRTTRLPSPTPRLPTSS PSAPVWLLSTLLATPVPTASLLGNLRPPSLLQGEVMGTPSSPRGPESPRLAAGPSPCWHLGAMHESRS RWTEPGCSQCWCEDGKVTCEKVRCEAACSHPIPSRDGGCCPSCTGCFHSGVVRAEGDVFSPPNENCTV CVCLAGNVSCISPECPSGPCQTPPQTDCCTCVPGRCYFHGRWYADGAVFSGGGDECTTCVCQNGEVEC SFMPCPELACPREEWRLGPGQCCFTCQEPTPSTGLDDNGVEFPIGQIWSPGDPCERWLGELQEDRLCG LLPSPDPDPWTVLPRLFSR		

	SEQ ID NO: 13	534 bp
NOV3d, 209749357 DNA Sequence	AAGCTTTGCTGGCACCTGGGAGCCATGCATGAATCAAGGAGTCGCTGGACAGAGCCTGGGTGTTCCC AGTGCTGGTGCGAGGACGGGAAGGTGACCTGTGAAAAGGTGAGGTGTGAAGCTGCTTGTTCCCACCC AATTCCCTCCAGAGATGGTGGGTGCTGCCCATCGTGCACAGGCTGTTTTCACAGTGGTGTCGTCCGA GCTGAAGGGGATGTGTTTTCACCTCCCAATGAGAACTGCACCGTCTGTGTCTGTCTGGCTGGAAACG TGTCCTGCATCTCTCCTGAGTGTCCTTCTGGCCCCTGTCAGACCCCCCCACAGACGGATTGCTGTAC TTGTGTTCCAGTGAGATGCTATTTCCACGGCCGGTGGTACGCAGACGGGGCTGTGTTCAGTGGGGGT GGTGACGAGTGTACCACCTGTGTTTGCCAGAATGGGGAGGTGGAGTGCTCCTTCATGCCCTGCCCTG AGCTGGCCTGCCCCCGAGAAGAGTGGCGGCTGGGCCCTGGGCAGTGTTGCTTCACCTGCCTCGAG	
	ORF Start: at 1	ORF Stop: end of sequence

	SEQ ID NO: 14	178 aa	MW at 19201.6kD
NOV3d, 209749357 Protein Sequence	KLCWHLGAMHESRSRWTEPGCSQCWCEDGKVTCEKVRCEAACSHPIPSRDGGCCPSCTGCFHSGVVR AEGDVFSPPNENCTVCVCLAGNVSCISPECPSGPCQTPPQTDCCTCVPVRCYFHGRWYADGAVFSGG GDECTTCVCQNGEVECSFMPCPELACPREEWRLGPGQCCFTCLE		

	SEQ ID NO: 15	534 bp	
NOV3e, CG106417-02 DNA Sequence	AAGCTTTGCTGGCACCTGGGAGCCATGCATGAATCAAGGAGTCGCTGGACAGAGCCTGGGTGTTCCCA GTGCTGGTGCGAGGACGGGAAGGTGACCTGTGAAAAGGTGAGGTGTGAAGCTGCTTGTTCCCACCCAA TTCCCTCCAGAGATGGTGGGTGCTGCCCATCGTGCACAGGCTGTTTTCACAGTGGTGTCGTCCGAGCT GAAGGGGATGTGTTTTCACCTCCCAATGAGAACTGCACCGTCTGTGTCTGTCTGGCTGGAAACGTGTC CTGCATCTCTCCTGAGTGTCCTTCTGGCCCCTGTCAGACCCCCCCACAGACGGATTGCTGTACTTGTG TTCCAGTGAGATGCTATTTCCACGGCCGGTGGTACGCAGACGGGGCTGTGTTCAGTGGGGGTGGTGAC GAGTGTACCACCTGTGTTTGCCAGAATGGGGAGGTGGAGTGCTCCTTCATGCCCTGCCCTGAGCTGGC CTGCCCCCGAGAAGAGTGGCGGCTGGGCCCTGGGCAGTGTTGCTTCACCTGCCTCGAG		
	ORF Start: at 7		ORF Stop: at 529

	SEQ ID NO: 16	174 aa	MW at 18718.0kD
NOV3e, CG106417-02 Protein Sequence	CWHLGAMHESRSRWTEPGCSQCWCEDGKVTCEKVRCEAACSHPIPSRDGGCCPSCTGCFHSGVVRAEG DVFSPPNENCTVCVCLAGNVSCISPECPSGPCQTPPQTDCCTCVPVRCYFHGRWYADGAVFSGGGDEC TTCVCQNGEVECSFMPCPELACPREEWRLGPGQCCFTC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 3B.

Table 3B. Comparison of NOV3a against NOV3b through NOV3e.		
Protein Sequence	**NOV3a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV3b	1..626 1..636	552/653 (84%) 552/653 (84%)
NOV3c	72..626 13..563	472/574 (82%) 475/574 (82%)
NOV3d	381..563 3..178	155/191 (81%) 156/191 (81%)
NOV3e	381..561 1..174	154/189 (81%) 155/189 (81%)

Further analysis of the NOV3a protein yielded the following properties shown in Table 3C.

Table 3C. Protein Sequence Properties NOV3a	
PSort analysis:	0.5947 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 22 and 23

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3D.

Table 3D. Geneseq Results for NOV3a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB85364	Novel Von Willebrand/thrombosporin-like polypeptide - Homo sapiens, 235 aa. [WO200153485-A1, 26-JUL-2001]	286..500 1..208	194/222 (87%) 196/222 (87%)	e-113
AAM99920	Human polypeptide SEQ ID NO 36 - Homo sapiens, 272 aa. [WO200155173-A2, 02-AUG-2001]	384..592 5..205	185/217 (85%) 188/217 (86%)	e-112
AAM99933	Human polypeptide SEQ ID NO 49 - Homo sapiens, 212 aa. [WO200155173-A2, 02-AUG-2001]	384..592 5..205	181/217 (83%) 185/217 (84%)	e-110
AAB85365	Novel Von Willebrand/thrombosporin-like mature protein sequence - Homo sapiens, 217 aa. [WO200153485-A1, 26-JUL-2001]	304..500 1..190	176/204 (86%) 178/204 (86%)	e-102
ABG15393	Novel human diagnostic protein #15384 - Homo sapiens, 1028 aa. [WO200175067-A2, 11-OCT-2001]	72..140 959..1027	69/69 (100%) 69/69 (100%)	8e-39

In a BLAST search of public sequence datbases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3E.

Table 3E. Public BLASTP Results for NOV3a

Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

Q96DN2	CDNA FLJ32009 fis, clone NT2RP7009498, weakly similar to fibulin-1, isoform A precursor - Homo sapiens (Human), 955 aa.	1..592 1..589	554/607 (91%) 558/607 (91%)	0.0
Q9DBE2	1300015B04Rik protein - Mus musculus (Mouse), 608 aa.	1..620 1..607	498/628 (79%) 530/628 (84%)	0.0
Q9NPY3	Complement component C1q receptor precursor (Complement component 1, q subcomponent, receptor 1) (C1qRp) (C1qR(p)) (C1q/MBL/SPA receptor) (CD93 antigen) (CDw93) - Homo sapiens (Human), 652 aa.	82..371 300..566	103/295 (34%) 132/295 (43%)	2e-32
Q9CXD8	6130401L20Rik protein - Mus musculus (Mouse), 528 aa.	54..260 96..305	78/219 (35%) 99/219 (44%)	2e-29
Q91V88	POEM (NEPHRONECTIN short isoform) - Mus musculus (Mouse), 561 aa.	45..368 35..383	100/363 (27%) 146/363 (39%)	3e-29

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3F.

Table 3F. Domain Analysis of NOV3a

Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value
EGF	148..181	16/47 (34%) 23/47 (49%)	0.0045
EGF	187..220	12/47 (26%) 25/47 (53%)	0.011
TIL	168..226	13/70 (19%) 39/70 (56%)	0.53
vwc	381..442	20/84 (24%) 41/84 (49%)	0.00069
vwc	452..502	18/84 (21%) 39/84 (46%)	0.00017
vwc	503..561	21/84 (25%) 39/84 (46%)	1.6e-05

Example 4.

The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis			
	SEQ ID NO: 17	1161 bp	
NOV4a, CG108901-01 DNA Sequence	GAATTCCGCAGCC**ATG**ACCCCGCAGCTTCTCCTGGCCCTTGTCCTCTGGCCAGCTGCCCGCCCTGCA GTGGAAGGAAAGGGCCCCCAGCAGCTCTGACACTGCCCCGGGTGCAATGCCGAGCCTCTCGGTACCCG ATCGCCGTGGATTGCTCCTGGACCCTGCCGCCTGCTCCAAACTCCACCAGCCCCGTGTCCTTCATTGC CACGTACAGGCTCGGCATGGCTGCCCGGGGCCACAGCTGGCCCTGCCTGCAGCAGACGCCAACGTCCA CCAGCTGCACCATCACGGATGTCCAGCTGTTCTCCATGGCTCCCTACGTGCTCAATGTCACCGCCGTC CACCCCTGGGGCTCCAGCAGCAGCTTCGTGCCTTTCATAACAGAGCACATCATCAAGCCCGACCCTCC AGAAGGCGTGCGCCTAAGCCCCCTCGCTGAGCGCCAGCTACAGGTGCAGTGGGAGCCTCCCGGGTCCT GGCCCTTCCCAGAGATCTTCTCACTGAAGTACTGGATCCGTTACAAGCGTCAGGGAGCTGCGCGCTTC CACCGGGTGGGGCCCATTGAAGCCACGTCCTTCATCCTCAGGGCTGTGCGGCCCCGAGCCAGGTACTA CGTCCAAGTGGCGGCTCAGGACCTCACAGACTACGGGGAACTGAGTGACTGGAGTCTCCCCGCCACTG CCACAATGAGCCTGGGCAAG**TAG**CAAGGGCTTCCCGCTGCCTCCAGACAGCACCTGGGTCCTCGCCAC CCTAAGCCCCGGGACACCTGTTGGAGGGCGGATGGATCTGCCTAGCCTGGGCTGGAGTCCTTGCTTT GCTGCTGCTGAGCTGCCGGGCAACCTCAGATGACCGACTTTTCCCTTTGAGCCTCAGTTTCTCTAGCT GAGAAATGGAGATGTACTACTCTCTCCTTTACCTTTACCTTTACCACAGTGCAGGGCTGACTGAACTG TCACTGTGAGATATTTTTTATTGTTTAATTAGAAAAGAATTGTTGTTGGCTGGGCGCAGTGGATCGC ACCTGTAATCCCAGTCACTGGGAAGCCGACGTGGGTGGGTAGCTTGAGGCCAGGAGCTCGAAACCAGT CCGGGCCACACAGCAAGACCCCATCTCTAAAAAATTAATATAAATATAAAATAAAAAAAAAAAAAAGG AATTC		
	ORF Start: ATG at 14		ORF Stop: TAG at 701
	SEQ ID NO: 18	229 aa	MW at 25396.0kD
NOV4a, CG108901-01 Protein Sequence	MTPQLLLALVLWASCPPCSGRKGPPAALTLPRVQCRASRYPIAVDCSWTLPPAPNSTSPVSFIATYRL GMAARGHSWPCLQQTPTSTSCTITDVQLFSMAPYVLNVTAVHPWGSSSSFVPFITEHIIKPDPPEGVR LSPLAERQLQVQWEPPGSWPFPEIFSLKYWIRYKRQGAARFHRVGPIEATSFILRAVRPRARYYVQVA AQDLTDYGELSDWSLPATATMSLGK		
	SEQ ID NO: 19	528 bp	
NOV4b, CG108901-04 DNA Sequence	CC**ATG**ACCCCGCAGCTTCTCCTGGCCCTTGTCCTCTGGCCAGCTGCCCGCCCTGCAGTGGAAGGAA AGGGCCCCCAGCAGCTCTGACACTGCCCCGGGTGCAATGCCGAGCCTCTCGGTACCCGATCGCCGTG GATTGCTCCTGGACCCTGCCGCCTGCTCCAAACTCCACCAGCCCCGTGCCTTTCATAACAGACCACA TCATCAAGCCCGACCCTCCAGAAGGCGTGCGCCTAAGCCCCCTCGCTGAGCGCCAGCTACAGGTGCA GTGGGAGCCTCCCGGGTCCTGGCCCTTCCCAGAGATCTTCTCACTGAAGTACTGGATCCGTTACAAG CGTCAGGGAGCTGCGCGCTTCCACCGGGTGGGGCCCATTGAAGCCACGTCCTTCATCCTCAGGGCTG TGCGGCCCCGAGCCAGGTACTACGTCCAAGTGGCGGCTCAGGACCTCACAGACTACGGGGAACTGAG TGACTGGAGTCTCCCCGCCACTGCCACAATGAGCCTGGGCAAG**TAG**CAAGGGCTTCCCG		
	ORF Start: ATG at 3		ORF Stop: TAG at 513
	SEQ ID NO: 20	170 aa	MW at 18991.8kD
NOV4b, CG108901-04 Protein Sequence	MTPQLLLALVLWASCPPCSGRKGPPAALTLPRVQCRASRYPIAVDCSWTLPPAPNSTSPVPFITDHI IKPDPPEGVRLSPLAERQLQVQWEPPGSWPFPEIFSLKYWIRYKRQGAARFHRVGPIEATSFILRAV RPRARYYVQVAAQDLTDYGELSDWSLPATATMSLGK		
	SEQ ID NO: 21	542 bp	
NOV4c,	C**ATG**ACCCCGCAGCTTCTCCTGGCCCTTGTCCTCTGGCCAGCTGCCCGCCCTGCAGTGGAAGGAAAG		

CG108901-03 DNA Sequence	GGCCCTGCCTGCAGCAGACGCCAACGTCCACCAGCTGCACCATCACGGATGTCCAGCTGTTCTCCATG GTTCCCTACGTGCTCAATGTCACCGCCGTCCACCCCTGGGGCTCCAGCAGCAGCTTCGTGCCTTTCAT AACAGAGCACATCATCAAGCCCGACCCTCCAGAAGGCGTGCGCCTAAGCCCCCTCGCTGAGCGCCAGC TACAGGTGCAGTGGGAGCCTCCTGGGTCCTGGCCCTTCCCAGAGATCTTCTCACTGAAGTACTGGATC CGTTACAAGCGTCAGGGAGCTGCGCGCTTCCACCGGGTGGGGCCCATTGAAGCCACGTCCTTCATCCT CAGGGCTGTGCGGCCCCGAGCCAGGTACTACATCCAAGTGGCGGCTCAGGACCTCACAGACTACGGGG AACTGAGTGACTGGAGTCTCCCCGCCACTGCCACAATGAGCCTGGGCAAGTAGCAAGGGCTTCCCG		
	ORF Start: ATG at 2		ORF Stop: TAG at 527

	SEQ ID NO: 22	175 aa	MW at 19616.5kD
NOV4c, CG108901-03 Protein Sequence	MTPQLLLALVLWASCPPCSGRKGPCLQQTPTSTSCTITDVQLFSMVPYVLNVTAVHPWGSSSSFVPFI TEHIIKPDPPEGVRLSPLAERQLQVQWEPPGSWPFPEIFSLKYWIRYKRQGAARFHRVGPIEATSFIL RAVRPRARYYIQVAAQDLTDYGELSDWSLPATATMSLGK		

	SEQ ID NO: 23	943 bp	
NOV4d, CG108901-02 DNA Sequence	CGGGAAGCCCTTGCTACTTGCCCAGGCTCATCGTGGCAGTGGCGGGGAGACTCCAGTCACTCAGTTC CCCGTAGTCTGTGAGGTCCTGAGCCGCCACTTGGATGTAGTACCTGGCTCGGGGCCGCACAGCCCTG AGGATGAAGGACGTGGCTTCAATGGGCCCCACCCGGTGGAAGCGCGCAGCTCCCTGACGCTTGTAAC GGATCCAGTACTTCAGTGAGAAGATCTCTGGGAAGGGCC**ATG**ACCCCGCAGCTTCTCCTGGCCCTTG TCCTCTGGGCCAGCTGCCCGCCCTGCAGTGGAAGGAAAGGGCCCCCAGCAGCTCTGACACTGCCCCG GGTGCAATGCCGAGCCTCTCGGTACCCGATCGCCGTGGATTGCTCCTGGACCCTGCCGCCTGCTCCA AACTCCACCAGCCCCGTGTCCTTCATTGCCACGTACAGGCTCGGCATGGCTGCCCGGGGCCACAGCT GGCCCTGCCTGCAGCAGACGCCAACGTCCACCAGCTGCACCATCACGGATGTCCAGCTGTTCTCCAT GGCTCCCTACGTGCTCAATGTCACCGCCGTCCACCCCTGGGGCTCCAGCAGCAGCTTCGTGCCTTTC ATAACAGAGCACATCATCAAGCCCGACCCTCCAGAAGGCGTGCGCCTAAGCCCCCTCGCTGAGCGCC AGCTACAGGTGCAGTGGGAGCCTCCCGGGTCCTGGCCCTTCCCAGAGATCTTCTCACTGAAGTACTG GATCCGTTACAAGCGTCAGGGAGCTGCGCGCTTCCACCGGGTGGGGCCCATTGAAGCCACGTCCTTC ATCCTCAGGGCTGTGCGGCCCCGAGCCAGGTACTACATCCAAGTGGCGGCTCAGGACCTCACAGACT ACGGGGAACTGAGTGACTGGAGTCTCCCCGCCACTGCCACGATGAGCCTGGGCAAG**TAG**CAAGGGCT TCCCG		
	ORF Start: ATG at 241		ORF Stop: TAG at 928

	SEQ ID NO: 24	229 aa	MW at 25410.0kD
NOV4d, CG108901-02 Protein Sequence	MTPQLLLALVLWASCPPCSGRKGPPAALTLPRVQCRASRYPIAVDCSWTLPPAPNSTSPVSFIATYR LGMAARGHSWPCLQQTPTSTSCTITDVQLFSMAPYVLNVTAVHPWGSSSSFVPFITEHIIKPDPPEG VRLSPLAERQLQVQWEPPGSWPFPEIFSLKYWIRYKRQGAARFHRVGPIEATSFILRAVRPRARYYI QVAAQDLTDYGELSDWSLPATATMSLGK		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 4B.

Table 4B. Comparison of NOV4a against NOV4b through NOV4d.

Protein Sequence	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV4b	1..229 1..170	156/229 (68%) 162/229 (70%)
NOV4c	1..229 1..175	170/229 (74%) 171/229 (74%)

NOV4d	1..229 1..229	228/229 (99%) 229/229 (99%)

Further analysis of the NOV4a protein yielded the following properties shown in Table 4C.

Table 4C. Protein Sequence Properties NOV4a	
PSort analysis:	0.8650 probability located in lysosome (lumen); 0.3700 probability located in outside; 0.1825 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 21 and 22

A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4D.

Table 4D. Geneseq Results for NOV4a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW09779	Epstein Barr virus-induced protein 3 (EBI3) - Homo sapiens, 229 aa. [WO9713859-A1, 17-APR-1997]	1..229 1..229	229/229 (100%) 229/229 (100%)	e-137
ABB81683	Human clone LO81-19a protein #1 - Homo sapiens, 229 aa. [WO200231114-A2, 18-APR-2002]	1..229 1..229	228/229 (99%) 229/229 (99%)	e-136
AAO14527	Human EBI-3 protein - Homo sapiens, 229 aa. [WO200212282-A2, 14-FEB-2002]	1..229 1..229	227/229 (99%) 228/229 (99%)	e-136
AAB36652	Human cytokine receptor subunit Eib3 protein SEQ ID NO:9 - Homo sapiens, 229 aa. [WO200073451-A1, 07-DEC-2000]	1..229 1..229	227/229 (99%) 228/229 (99%)	e-136
AAW53624	Epstein Barr virus induced gene 3 (EBI-3) - Homo sapiens, 229 aa. [US5744301-A,	1..229 1..229	227/229 (99%) 228/229 (99%)	e-136

	28-APR-1998]			

In a BLAST search of public sequence datbases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4E.

Table 4E. Public BLASTP Results for NOV4a

Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O75269	Human cytokine receptor (Epstein-Barr virus induced gene 3) - Homo sapiens (Human), 229 aa.	1..229 1..229	229/229 (100%) 229/229 (100%)	e-136
Q14213	Cytokine receptor precursor - Homo sapiens (Human), 229 aa.	1..229 1..229	227/229 (99%) 228/229 (99%)	e-135
O35228	Cytokine receptor-like molecule (Epstein-Barr virus induced gene 3) - Mus musculus (Mouse), 228 aa.	1..220 1..218	138/220 (62%) 166/220 (74%)	5e-75
CAD29041	Sequence 29 from Patent WO0214358 - Homo sapiens (Human), 102 aa.	1..67 1..67	67/67 (100%) 67/67 (100%)	3e-34
CAD44518	SI:bZ76A6.1 (novel protein similar to vertebrate ciliary neurotrophic factor receptor alpha (CNTFR alpha)) - Brachydanio rerio (Zebrafish) (Danio rerio), 212 aa (fragment).	31..224 5..193	65/196 (33%) 99/196 (50%)	5e-24

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4F.

Table 4F. Domain Analysis of NOV4a

Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
fn3	129..215	19/89 (21%) 56/89 (63%)	0.0001

Example 5.

The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 5A.

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 25	3971 bp	
NOV5a, CG112505-01 DNA Sequence	GCTTTCAGGCGATCTGGAGAAAGAACGGCAGAACACACAGCAAGGAAAGGTCCTTTCTGGGGATCACC CCATTGGCTGAAGATGAGACCATTCTTCCTCTTGTGTTTTGCCCTGCCTGGCCTCCTGCATGCCCAAC AAGCCTGCTCCCGTGGGGCCTGCTATCCACCTGTTGGGGACCTGCTTGTTGGGAGGACCCGGTTTCTC CGAGCTTCATCTACCTGTGGACTGACCAAGCCTGAGACCTACTGCACCCAGTATGGCGAGTGGCAGAT GAAATGCTGCAAGTGTGACTCCAGGCAGCCTCACAACTACTACAGTCACCGAGTAGAGAATGTGGCTT CATCCTCCGGCCCCATGCGCTGGTGGCAGTCCCAGAATGATGTGAACCCTGTCTCTCTGCAGCTGGAC CTGGACAGGAGATTCCAGCTTCAAGAAGTCATGATGGAGTTCCAGGGGCCCATGCCCGCCGGCATGCT GATTGAGCGCTCCTCAGACTTCGGTAAGACCTGGCGAGTGTACCAGTACCTGGCTGCCGACTGCACCT CCACCTTCCCTCGGGTCCGCCAGGGTCGGCCTCAGAGCTGGCAGGATGTTCGGTGCCAGTCCCTGCCT CAGAGGCCTAATGCACGCCTAAATGGGGGGAAGGTCCAACTTAACCTTATGGATTTAGTGTCTGGGAT TCCAGCAACTCAAAGTCAAAAAATTCAAGAGGTGGGGGAGATCACAAACTTGAGAGTCAATTTCACCA GGCTGGCCCCTGTGCCCCAAAGGGGCTACCACCCTCCCAGCGCCTACTATGCTGTGTCCCAGCTCCGT CTGCAGGGGAGCTGCTTCTGTCACGGCCATGCTGATCGCTGCGCACCCAAGCCTGGGGCCTCTGCAGG CCCCTCCACCGCTGTGCAGGTCCACGATGTCTGTGTCTGCCAGCACAACACTGCCGGCCCAAATTGTG AGCGCTGTGCACCCTTCTACAACAACCGGCCCTGGAGACCGGCGGAGGGCCAGGACGCCCATGAATGC CAAAGGTGCGACTGCAATGGGCACTCAGAGACATGTCACTTTGACCCCGCTGTGTTTGCCGCCAGCCA GGGGGCATATGGAGGTGTGTGTGACAATTGCCGGGACCACACCGAAGGCAAGAACTGTGAGCGGTGTC AGCTGCACTATTTCCGGAACCGGCGCCCGGGAGCTTCCATTCAGGAGACCTGCATCTCCTGCGAGTGT GATCCGGATGGGGCAGTGCCAGGGGCTCCCTGTGACCCAGTGACCGGGCAGTGTGTGTGCAAGGAGCA TGTGCAGGGAGAGCGCTGTGACCTATGCAAGCCGGGCTTCACTGGACTCACCTACGCCAACCCGCAGG GCTGCCACCGCTGTGACTGCAACATCCTGGGGTCCCGGAGGGACATGCCGTGTGACGAGGAGAGTGGG CGCTGCCTTTGTCTGCCCAACGTGGTGGGTCCCAAATGTGACCAGTGTGCTCCCTACCACTGGAAGCT GGCCAGTGGCCAGGGCTGTGAACCGTGTGCCTGCGACCCGCACAACTCCCCTCAGCCCACAGTGCAAC CAGTTCACAGGGCAGTGCCCTGTCGGGAAGGCTTTGGTGGCCTGATGTGCAGCGCTGCAGCCATCCGC CAGTGTCCAGACCGGACCTATGGAGACGTGGCCACAGGATGCCGAGCCTGTGACTGTGATTTCCGGGG AACAGAGGGCCCGGGCTGCGACAAGGCATCAGGCCGCTGCCTCTGCCGCCCTGGCTTGACCGGGCCCC GCTGTGACCAGTGCCAGCGAGGCTACTGCAATCGCTACCCGGTGTGCGTGGCCTGCCACCCTTGCTTC CAGACCTATGATGCGGACCTCCGGGAGCAGGCCCTGCGCTTTGGTAGACTCCGCAATGCCACCGCCAG CCTGTGGTCAGGGCCTGGGCTGGAGGACCGTGGCCTGGCCTCCCGGATCCTAGATGCAAAGAGTAAGA TTGAGCAGATCCGAGCAGTTCTCAGCAGCCCCGCAGTCACAGAGCAGGAGGTGGCTCAGGTGGCCAGT GCCATCCTCTCCCTCAGGCGAACTCTCCAGGGCCTGCAGCTGGATCTGCCCCTGGAGGAGGAGACGTT GTCCCTTCCGAGAGACCTGGAGAGTCTTGACAGAAGCTTCAATGGTCTCCTTACTATGTATCAGAGGA AGAGGGAGCAGTTTGAAAAAATAAGCAGTGCTGATCCTTCAGGAGCCTTCCGGATGCTGAGCACAGCC TACGAGCAGTCAGCCCAGGCTGCTCAGCAGGTCTCCGACAGCTCGCGCCTTTTGGACCAGCTCAGGGA CAGCCGGAGAGAGGCAGAGAGGCTGGTGCGGCAGGCGGGAGGAGGAGGAGGCACCGGCAGCCCCAAGC TTGTGGCCCTGAGGCTGGAGATGTCTTCGTTGCCTGACCTGACACCCACCTTCAACAAGCTCTGTGGC AACTCCAGGCAGATGGCTTGCACCCCAATATCATGCCCTGGTGAGCTATGTCCCCAAGACAATGGCAC AGCCTGTGGCTCCCGCTGCAGGGGTGTCCTTCCCAGGGCCGGTGGGGCCTTCTTGATGGCGGGGCAGG TGGCTGAGCAGCTGCGGGGCTTCAATGCCCAGCTCCAGCGGACCAGGCAGATGATTAGGGCAGCCGAG GAATCTGCCTCACAGATTCAATCCAGTGCCCAGCGCTTGGAGACCCAGGTGAGCGCCAGCCGCTCCCA GATGGAGGAAGATGTCAGACGCACACGGCTCCTAATCCAGCAGGTCCGGGACTTCCTAACAGACCCCG ACACTGATGCAGCCACTATCCAGGAGGTCAGCGAGGCCGTGCTGGCCCTGTGGCTGCCCACAGACTCA GCTACTGTTCTGCAGAAGATGAATGAGATCCAGGCCATTGCAGCCAGGCTCCCCAACGTGGACTTGGT GCTGTCCCAGACCAAGCAGGACATTGCGCGTGCCCGCCGGTTGCAGGCTGAGGCTGAGGAAGCCAGGA GCCGAGCCCATGCAGTGGAGGGCCAGGTGGAAGATGTGGTTGGGAACCTGCGGCAGGGGACAGTGGCA CTGCAGGAAGCTCAGGACACCATGCAAGGCACCAGCCGCTCCCTTCGGCTTATCCAGGACAGGGTTGC TGAGGTTCAGCAGGTACTGCGGCCAGCAGAAAAGCTGGTGACAAGCATGACCAAGCAGCTGGGTGACT TCTGGACACGGATGGAGGAGCTCCGCCACCAAGCCCGGCAGCAGGGGGCAGAGGCAGTCCAGGCCCAG CAGCTTGCGGAAGGTGCCAGCGAGCAGGCATTGAGTGCCCAAGAGGGATTTGAGAGAATAAAACAAAA GTATGCTGAGTTGAAGGACCGGTTGGGTCAGAGTTCCATGCTGGGTGAGCAGGGTGCCCGGATCCAGA GTGTGAAGACAGAGGCAGAGGAGCTGTTTGGGGAGACCATGGAGATGATGGACAGGATGAAAGACATG GAGTTGGAGCTGCTGCGGGGCAGCCAGGCCATCATGCTGCGCTCGGCGGACCTGACAGGACTGGAGAA GCGTGTGGAGCAGATCCGTGACCACATCAATGGGCGCGTGCTCTACTATGCCACCTGCAAGTGATGCT ACAGCTTCCAGCCCGTTGCCCCACTCATCTGCCGCCTTTGCTTTTGGTTGGGGGCAGATTGGGTTGGA ATGCTTTCCATCTCCAGGAGACTTTCATGCAGCCTAAAGTACAGCCTGGACCACCCCTGGTGTGTAGC TAGTAAGATTACCCTGAGCTGCAGCTGAGCCTGAGCCAATGGACAGTTACACTTGACAGACAAAGAT GGTGGAGATTGGCATGCCATTGAAACTAAGAGCTCTCAAGTCAAGGAAGCTGGGCTGGGCAGTATCCC CCGCCTTTAGTTCTCCACTGGGGAGGAATCCTGGACCAAGCACAAAAACTTAACAAAAGTGATGTAAA AATGAAAAGCCAAATAAAAATCTTTGG		
	ORF Start: ATG at 82		ORF Stop: TGA at 3598

	SEQ ID NO: 26	1172 aa	MW at 129574.1kD
NOV5a, CG112505-01 Protein Sequence	MRPFFLLCFALPGLLHAQQACSRGACYPPVGDLLVGRTRFLRASSTCGLTKPETYCTQYGEWQMKCCK CDSRQPHNYYSHRVENVASSSGPMRWWQSQNDVNPVSLQLDLDRRFQLQEVMMEFQGPMPAGMLIERS SDFGKTWRVYQYLAADCTSTFPRVRQGRPQSWQDVRCQSLPQRPNARLNGGKVQLNLMDLVSGIPATQ SQKIQEVGEITNLRVNFTRLAPVPQRGYHPPSAYYAVSQLRLQGSCFCHGHADRCAPKPGASAGPSTA VQVHDVCVCQHNTAGPNCERCAPFYNNRPWRPAEGQDAHECQRCDCNGHSETCHFDPAVFAASQGAYG GVCDNCRDHTEGKNCERCQLHYFRNRRPGASIQETCISCECDPDGAVPGAPCDPVTGQCVCKEHVQGE RCDLCKPGFTGLTYANPQGCHRCDCNILGSRRDMPCDEESGRCLCLPNVVGPKCDQCAPYHWKLASGQ GCEPCACDPHNSPQPTVQPVHRAVPCREGFGGLMCSAAAIRQCPDRTYGDVATGCRACDCDFRGTEGP GCDKASGRCLCRPGLTGPRCDQCQRGYCNRYPVCVACHPCFQTYDADLREQALRFGRLRNATASLWSG PGLEDRGLASRILDAKSKIEQIRAVLSSPAVTEQEVAQVASAILSLRRTLQGLQLDLPLEEETLSLPR DLESLDRSFNGLLTMYQRKREQFEKISSADPSGAFRMLSTAYEQSAQAAQQVSDSSRLLDQLRDSRRE AERLVRQAGGGGGTGSPKLVALRLEMSSLPDLTPTFNKLCGNSRQMACTPISCPGELCPQDNGTACGS RCRGVLPRAGGAFLMAGQVAEQLRGFNAQLQRTRQMIRAAEESASQIQSSAQRLETQVSASRSQMEED VRRTRLLIQQVRDFLTDPDTDAATIQEVSEAVLALWLPTDSATVLQKMNEIQAIAARLPNVDLVLSQT KQDIARARRLQAEAEEARSRAHAVEGQVEDVVGNLRQGTVALQEAQDTMQGTSRSLRLIQDRVAEVQQ VLRPAEKLVTSMTKQLGDFWTRMEELRHQARQQGAEAVQAQQLAEGASEQALSAQEGFERIKQKYAEL KDRLGQSSMLGEQGARIQSVKTEAEELFGETMEMMDRMKDMELELLRGSQAIMLRSADLTGLEKRVEQ IRDHINGRVLYYATCK		
	SEQ ID NO: 27	3810 bp	
NOV5b, CG112505-02 DNA Sequence	GCTTTCAGGCGATCTGGAGAAAGAACGGCAGAACACACAGCAAGGAAAGGTCCTTTCTGGGGATCAC CCCATTGGCTGAAG**ATG**AGACCATTCTTCCTCTTGTGTTTTGCCCTGCCTGGCCTCCTGCATGCCCA ACAAGCCTGCTCCCGTGGGGCCTGCTATCCACCTGTTGGGGACCTGCTTGTTGGGAGGACCCGGTTT CTCCGAGCTTCATCTACCTGTGGACTGACCAAGCCTGAGACCTACTGCACCCAGTATGGCGAGTGGC AGATGAAATGCTGCAAGTGTGACTCCAGGCAGCCTCACAACTACTACAGTCACCGAGTAGAGAATGT GGCTTCATCCTCCGGCCCCATGCGCTGGTGGCAGTCCCAGAATGATGTGAACCCTGTCTCTCTGCAG CTGGACCTGGACAGGAGATTCCAGCTTCAAGAAGTCATGATGGAGTTCCAGGGGCCCATGCCCGCCG GCATGCTGATTGAGCGCTCCTCAGACTTCGGTAAGACCTGGCGAGTGTACCAGTACCTGGCTGCCGA CTGCACCTCCACCTTCCCTCGGGTCCGCCAGGGTCGGCCTCAGAGCTGGCAGGATGTTCGGTGCCAG TCCCTGCCTCAGAGGCCTAATGCACGCCTAAATGGGGGGAAGGTCCAACTTAACCTTATGGATTTAG TGTCTGGGATTCCAGCAACTCAAAGTCAAAAAATTCAAGAGGTGGGGGAGATCACAAACTTGAGAGT CAATTTCACCAGGCTGGCCCCTGTGCCCCAAAGGGGCTACCACCCTCCCAGCGCCTACTATGCTGTG TCCCAGCTCCGTCTGCAGGGGAGCTGCTTCTGTCACGGCCATGCTGATCGCTGCGCACCCAAGCCTG GGGCCTCTGCAGGCCCCTCCACCGCTGTGCAGGTCCACGATGTCTGTGTCTGCCAGCACAACACTGC CGGCCCAAATTGTGAGCGCTGTGCACCCTTCTACAACAACCGGCCCTGGAGACCGGCGGAGGGCCAG GACGCCCATGAATGCCAAAGGTGCGACTGCAATGGGCACTCAGAGACATGTCACTTTGACCCCGCTG TGTTTGCCGCCAGCCAGGGGGCATATGGAGGTGTGTGTGACAATTGCCGGGACCACACCGAAGGCAA GAACTGTGAGCGGTGTCAGCTGCACTATTTCCGGAACCGGCGCCCGGGAGCTTCCATTCAGGAGACC TGCATCTCCTGCGAGTGTGATCCGGATGGGGCAGTGCCAGGGGCTCCCTGTGACCCAGTGACCGGGC AGTGTGTGTGCAAGGAGCATGTGCAGGGAGAGCGCTGTGACCTATGCAAGCCGGGCTTCACTGGACT CACCTACGCCAACCCGCAGGGCTGCCACCGCTGTGACTGCAACATCCTGGGGTCCCGGAGGGACATG CCGTGTGACGAGGAGAGTGGGCGCTGCCTTTGTCTGCCCAACGTGGTGGGTCCCAAATGTGACCAGT GTGCTCCCTACCACTGGAAGCTGGCCAGTGGCCAGGGCTGTGAACCGTGTGCCTGCGACCCGCACAA CTCCCCTCAGCCCACAGTGCAACCAGTTCACAGGGCAGTGCCCTGTCGGGAAGGCTTTGGTGGCCTG ATGTGCAGCGCTGCAGCCATCCGCCAGTGTCCAGACCGGACCTATGGAGACGTGGCCACAGGATGCC GAGCCTGTGACTGTGATTTCCGGGGAACAGAGGGCCCGGGCTGCGACAAGGCATCAGGCCGCTGCCT CTGCCGCCCTGGCTTGACCGGGCCCCGCTGTGACCAGTGCCAGCGAGGCTACTGCAATCGCTACCCG GTGTGCGTGGCCTGCCACCCTTGCTTCCAGACCTATGATGCGGACCTCCGGGAGCAGGCCCTGCGCT TTGGTAGACTCCGCAATGCCACCGCCAGCCTGTGGTCAGGGCCTGGGCTGGAGGACCGTGGCCTGGC CTCCCGGATCCTAGATGCAAAGAGTAAGATTGAGCAGATCCGAGCAGTTCTCAGCAGCCCCGCAGTC ACAGAGCAGGAGGTGGCTCAGGTGGCCAGTGCCATCCTCTCCCTCAGGAGCCTTCCGGATGCTGAGC ACAGCCTACGAGCAGTCAGCCCAGGCTGCTCAGCAGGTCTCCGACAGCTCGCGCCTTTTGGACCAGC TCAGGGACAGCCGGAGAGAGGCAGAGAGGCTGGTGCGGCAGGCGGGAGGAGGAGGAGGCACCGGCAG CCCCAAGCTTGTGGCCCTGAGGCTGGAGATGTCTTCGTTGCCT**GA**CCTGACACCCACCTTCAACAAG CTCTGTGGCAACTCCAGGCAGATGGCTTGCACCCCAATATCATGCCCTGGTGAGCTATGTCCCCAAG ACAATGGCACAGCCTGTGGCTCCCGCTGCAGGGGTGTCCTTCCCAGGGCCGGTGGGGCCTTCTTGAT GGCGGGGCAGGTGGCTGAGCAGCTGCGGGGCTTCAATGCCCAGCTCCAGCGGACCAGGCAGATGATT AGGGCAGCCGAGGAATCTGCCTCACAGATTCAATCCAGTGCCCAGCGCTTGGAGACCCAGGTGAGCG CCAGCCGCTCCCAGATGGAGGAAGATGTCAGACGCACACGGCTCCTAATCCAGCAGGTCCGGGACTT CCTAACAGACCCCGACACTGATGCAGCCACTATCCAGGAGGTCAGCGAGGCCGTGCTGGCCCTGTGG CTGCCCACAGACTCAGCTACTGTTCTGCAGAAGATGAATGAGATCCAGGCCATTGCAGCCAGGCTCC CCAACGTGGACTTGGTGCTGTCCCAGACCAAGCAGGACATTGCGCGTGCCCGCCGGTTGCAGGCTGA GGCTGAGGAAGCCAGGAGCCGAGCCCATGCAGTGGAGGGCCAGGTGGAAGATGTGGTTGGGAACCTG CGGCAGGGGACAGTGGCACTGCAGGAAGCTCAGGACACCATGCAAGGCACCAGCCGCTCCCTTCGGC TTATCCAGGACAGGGTTGCTGAGGTTCAGCAGGTACTGCGGCCAGCAGAAAAGCTGGTGACAAGCAT GACCAAGCAGCTGGGTGACTTCTGGACACGGATGGAGGAGCTCCGCCACCAAGCCCGGCAGCAGGGG GCAGAGGCAGTCCAGGCCCAGCAGCTTGCGGAAGGTGCCAGCGAGCAGGCATTGAGTGCCCAAGAGG GATTTGAGAGAATAAAACAAAAGTATGCTGAGTTGAAGGACCGGTTGGGTCAGAGTTCCATGCTGGG TGAGCAGGGTGCCCGGATCCAGAGTGTGAAGACAGAGGCAGAGGAGCTGTTTGGGGAGACCATGGAG		

	ATGATGGACAGGATGAAAGACATGGAGTTGGAGCTGCTGCGGGGCAGCCAGGCCATCATGCTGCGCT CGGCGGACCTGACAGGACTGGAGAAGCGTGTGGAGCAGATCCGTGACCACATCAATGGGCGCGTGCT CTACTATGCCACCTGCAAGTGATGCTACAGCTTCCAGCCCGTTGCCCCACTCATCTGCCGCCTTTGC TTTTGGTTGGGGGCAGATTGGGTTGGAATGCTTTCCATCTCCAGGAGACTTTCATGCAGCCTAAAGT ACAGCCTGGACCACCCCTGGTGTGTAGCTAGTAAGATTACCCTGAGCTGCAGCTGAGCCTGAGCCAA TGGGACAGTTACACTTGACAGACAAAGATGGTGGAGATTGGCATGCCATTGAAACTAAGAGCTCTCA AGTCAAGGAAGCTGGGCTGGGCAGTATCCCCCGCCTTTAGTTCTCCACTGGGGAGGAATCCTGGACC AAGCACAAAAACTTAACAAAAGTGATGTAAAAATGAAAAGCCAAATAAAAATCTTTGG		
	ORF Start: ATG at 82		ORF Stop: TGA at 2254

	SEQ ID NO: 28	724 aa	MW at 79264.7kD
NOV5b, CG112505-02 Protein Sequence	MRPFFLLCFALPGLLHAQQACSRGACYPPVGDLLVGRTRFLRASSTCGLTKPETYCTQYGEWQMKCC KCDSRQPHNYYSHRVENVASSSGPMRWWQSQNDVNPVSLQLDLDRRFQLQEVMMEFQGPMPAGMLIE RSSDFGKTWRVYQYLAADCTSTFPRVRQGRPQSWQDVRCQSLPQRPNARLNGGKVQLNLMDLVSGIP ATQSQKIQEVGEITNLRVNFTRLAPVPQRGYHPPSAYYAVSQLRLQGSCFCHGHADRCAPKPGASAG PSTAVQVHDVCVCQHNTAGPNCERCAPFYNNRPWRPAEGQDAHECQRCDCNGHSETCHFDPAVFAAS QGAYGGVCDNCRDHTEGKNCERCQLHYFRNRRPGASIQETCISCECDPDGAVPGAPCDPVTGQCVCK EHVQGERCDLCKPGFTGLTYANPQGCHRCDCNILGSRRDMPCDEESGRCLCLPNVVGPKCDQCAPYH WKLASGQGCEPCACDPHNSPQPTVQPVHRAVPCREGFGGLMCSAAAIRQCPDRTYGDVATGCRACDC DFRGTEGPGCDKASGRCLCRPGLTGPRCDQCQRGYCNRYPVCVACHPCFQTYDADLREQALRFGRLR NATASLWSGPGLEDRGLASRILDAKSKIEQIRAVLSSPAVTEQEVAQVASAILSLRSLPDAEHSLRA VSPGCSAGLRQLAPFGPAQGQPERGREAGAAGGRRRRHRQPQACGPEAGDVFVA		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 5B.

Table 5B. Comparison of NOV5a against NOV5b.		
Protein Sequence	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV5b	1..659 1..659	647/659 (98%) 647/659 (98%)

Further analysis of the NOV5a protein yielded the following properties shown in Table 5C.

Table 5C. Protein Sequence Properties NOV5a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 18 and 19

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 5D.

Table 5D. Geneseq Results for NOV5a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW37870	Human protein comprising secretory signal amino acid sequence 7 - Homo sapiens, 1172 aa. [WO9811217-A2, 19-MAR-1998]	1..1172 1..1172	1161/1172 (99%) 1161/1172 (99%)	0.0
AAB48466	Human laminin 5 polypeptide, SEQ ID NO: 22 - Homo sapiens, 1174 aa. [WO200066731-A2, 09-NOV-2000]	4..1172 6..1174	1151/1169 (98%) 1151/1169 (98%)	0.0
AAB48462	Human laminin 5 polypeptide, SEQ ID NO: 14 - Homo sapiens, 1170 aa. [WO200066731-A2, 09-NOV-2000]	1..1172 1..1170	1152/1172 (98%) 1155/1172 (98%)	0.0
AAB48464	Human laminin 5 polypeptide, SEQ ID NO: 18 - Homo sapiens, 1186 aa. [WO200066731-A2, 09-NOV-2000]	4..1172 6..1186	1152/1181 (97%) 1152/1181 (97%)	0.0
AAB48465	Human laminin 5 polypeptide, SEQ ID NO: 20 - Homo sapiens, 1167 aa. [WO200066731-A2, 09-NOV-2000]	17..1172 12..1167	1145/1156 (99%) 1145/1156 (99%)	0.0

In a BLAST search of public sequence datbases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5E.

Table 5E. Public BLASTP Results for NOV5a				
Protein Accession Number	**Protein/Organism/Length**	**NOV5a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q13751	Laminin beta-3 chain precursor (Laminin 5 beta 3) (Laminin B1k chain) (Kalinin B1 chain) - Homo sapiens (Human), 1172 aa.	1..1172 1..1172	1161/1172 (99%) 1161/1172 (99%)	0.0

CAC17363	Sequence 21 from Patent WO0066731 precursor - Homo sapiens (Human), 1174 aa.	4..1172 6..1174	1151/1169 (98%) 1151/1169 (98%)	0.0
CAC17359	Sequence 13 from Patent WO0066731 precursor - Homo sapiens (Human), 1170 aa.	1..1172 1..1170	1152/1172 (98%) 1155/1172 (98%)	0.0
CAC17361	Sequence 17 from Patent WO0066731 precursor - Homo sapiens (Human), 1186 aa.	4..1172 6..1186	1152/1181 (97%) 1152/1181 (97%)	0.0
CAC17362	Sequence 19 from Patent WO0066731 - Homo sapiens (Human), 1167 aa (fragment).	17..1172 12..1167	1145/1156 (99%) 1145/1156 (99%)	0.0

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5F.

Table 5F. Domain Analysis of NOV5a

Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value
laminin_Nterm	26..248	88/273 (32%) 150/273 (55%)	1.6e-38
laminin_EGF	250..313	17/70 (24%) 50/70 (71%)	4e-08
laminin_EGF	316..376	19/65 (29%) 50/65 (77%)	1.7e-13
laminin_EGF	379..428	26/59 (44%) 43/59 (73%)	9.4e-18
laminin_EGF	431..478	27/59 (46%) 39/59 (66%)	3.9e-17
laminin_EGF	481..531	14/64 (22%) 34/64 (53%)	0.79
laminin_EGF	534..578	20/59 (34%) 34/59 (58%)	3.1e-10

Example 6.

The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis

	SEQ ID NO: 29	2659 bp	
NOV6a, CG121965-01 DNA Sequence	ACCCACGGGGCTGCCCTCCCCTGCGCACTCCCCTCGCTGCCCGGGCCCGGAGCGCAGTGGGGCCGCAC AGATTCACAATGTTGAAAGCCCTTTTCCTAACTATGCTGACTCTGGCGCTGGTCAAGTCACAGGACAC CGAAGAAACCATCACGTACACGCAATGCACTGACGGATATGAGTGGGATCCTGTGAGACAGCAATGCA AAGATATTGATGAATGTGACATTGTCCCAGACGCTTGTAAAGGTGGAATGAAGTGTGTCAACCACTAT GGAGGATACCTCTGCCTTCCGAAAACAGCCCAGATTATTGTCAATAATGAACAGCCTCAGCAGGAAAC ACAACCAGCAGAAGGAACCTCAGGGGCAACCACCGGGGTTGTAGCTGCCAGCAGCATGGCAACCAGTG GAGTGTTGCCCGGGGGTGGTTTTGTGGCCAGTGCTGCTGCAGTCGCAGGCCCTGAAATGCAGACTGGC CGAAATAACTTTGTCATCCGGCGGAACCCAGCTGACCCTCAGCGCATTCCCTCCAACCCTTCCCACCG TATCCAGTGTGCAGCAGGCTACGAGCAAAGTGAACACAACGTGTGCCAAGACATAGACGAGTGCACTG CAGGGACGCACAACTGTAGAGCAGACCAAGTGTGCATCAATTTACGGGGATCCTTTGCATGTCAGTGC CCTCCTGGATATCAGAAGCGAGGGGAGCAGTGCGTAGACATAGATGAATGTACCATCCCTCCATATTG CCACCAAAGATGCGTGAATACACCAGGCTCATTTTATTGCCAGTGCAGTCCTGGGTTTCAATTGGCAG CAAACAACTATACCTGCGTAGATATAAATGAATGTGATGCCAGCAATCAATGTGCTCAGCAGTGCTAC AACATTCTTGGTTCATTCATCTGTCAGTGCAATCAAGGATATGAGCTAAGCAGTGACAGGCTCAACTG TGAAGACATTGATGAATGCAGAACCTCAAGCTACCTGTGTCAATATCAATGTGTCAATGAACCTGGGA AATTCTCATGTATGTGCCCCCAGGGATACCAAGTGGTGAGAAGTAGAACATGTCAAGATATAAATGAG TGTGAGACCACAAATGAATGCCGGGAGGATGAAATGTGTTGGAATTATCATGGCGGCTTCCGTTGTTA TCCACGAAATCCTTGTCAAGATCCCTACATTCTAACACCAGAGAACCGATGTGTTTGCCCAGTCTCAA ATGCCATGTGCCGAGAACTGCCCCAGTCAATAGTCTACAAATACATGAGCATCCGATCTGATAGGTCT GTGCCATCAGACATCTTCCAGATACAGGCCACAACTATTTATGCCAACACCATCAATACTTTTCGGAT TAAATCTGGAAATGAAAATGGAGAGTTCTACCTACGACAAACAAGTCCTGTAAGTGCAATGCTTGTGC TCGTGAAGTCATTATCAGGACCAAGAGAACATATCGTGGACCTGGAGATGCTGACAGTCAGCAGTATA GGGACCTTCCGCACAAGCTCTGTGTTAAGATTGACAATAATAGTGGGGCCATTTTCATTTTAGTCTTT TCTAAGAGTCAACCACAGGCATTTAAGTCAGCCAAAGAATATTGTTACCTTAAAGCACTATTTTATTT ATAGATATATCTAGTGCATCTACATCTCTATACTGTACACTCACCCATAATTCAAACAATTACACCAT GGTATAAAGTGGGCATTTAATATGTAAAGATTCAAAGTTTGTCTTTATTACTATATGTAAATTAGACA TTAATCCACTAAACTGGTCTTCTTCAAGAGAGCTAAGTATACACTATCTGGTGAAACTTGGATTCTTT CCTATAAAAGTGGACCAAGCAATGATGATCTTCTGTGGTGCTTAAGGAAACTTACTAGAGCTCCACT AACAGTCTCATAAGGAGGCAGCCATCATAACCATTGAATAGCATGCAAGGGTAAGAATGAGTTTTTAA CTGCTTTGTAAGAAAATGGAAAAGGTCAATAAAGATATATTTCTTTAGAAAATGGGGATCTGCCATAT TTGTGTTGGTTTTTATTTTCATATCCAGCCTAAAGGTGGTTGTTTATTATATAGTAATAAATCATTGC TGTACAATATGCTGGTTTCTGTAGGGTATTTTTAATTTTGTCAGAAATTTTAGATTGTGAATATTTTG TAAAAAACAGTAAGCAAAATTTTCCAGAATTCCCAAAATGAACCAGATATCCCCTAGAAAATTATACT ATTGAGAAATCTATGGGGAGGATATGAGAAAATAAATTCCTTCTAAACCACATTGGAACTGACCTGAA GAAGCAAACTCGGAAAATATAATAACATCCCTGAATTCAGGACTTCCACAAGATGCAGAACAAAATGG ATAAAAGGTATTTCACTGGAGAAGTTTTAATTTCTAAGTAAAATTTAAATCCTAACACTTCACTAATT TATAACTAAAATTTCTCATCTTCGTACTTGATGCTCACAGAGGAAGAAAATGATGATGGTTTTTATTC CTGGCATCCAGAGTGACAGTGAACTTAAGCAAATTACCCTCCTACCCAATTCTATGGAATATTTTATA CGTCTCCTTGTTTAAAATGTCACTGCTTTACTTTGATGTATCATATTTTTAAATAAAAATAAATATTC CTTTAGA		
	ORF Start: ATG at 78		ORF Stop: TAG at 1557

	SEQ ID NO: 30	493 aa	MW at 54640.0kD
NOV6a, CG121965-01 Protein Sequence	MLKALFLTMLTLALVKSQDTEETITYTQCTDGYEWDPVRQQCKDIDECDIVPDACKGGMKCVNHYGGY LCLPKTAQIIVNNEQPQQETQPAEGTSGATTGVVAASSMATSGVLPGGGFVASAAAVAGPEMQTGRNN FVIRRNPADPQRIPSNPSHRIQCAAGYEQSEHNVCQDIDECTAGTHNCRADQVCINLRGSFACQCPPG YQKRGEQCVDIDECTIPPYCHQRCVNTPGSFYCQCSPGFQLAANNYTCVDINECDASNQCAQQCYNIL GSFICQCNQGYELSSDRLNCEDIDECRTSSYLCQYQCVNEPGKFSCMCPQGYQVVRSRTCQDINECET TNECREDEMCWNYHGGFRCYPRNPCQDPYILTPENRCVCPVSNAMCRELPQSIVYKYMSIRSDRSVPS DIFQIQATTIYANTINTFRIKSGNENGEFYLRQTSPVSAMLVLVKSLSGPREHIVDLEMLTVSSIGTF RTSSVLRLTIIVGPFSF		

	SEQ ID NO: 31	2625 bp	
NOV6b, CG121965-02	CTAGTATTCTACTAGAACTGGAAGATTGCTCTCCGAGTTTTGTTTTGTTATTTTGTTTAAAAAATAA AAAGCTTGAGGCCAAGGCAATTCATATTGGCTCACAGGTATTTTTGCTGTGCTGTGCAAGGAACTCT GCTAGCTCAAGATTCACAATGTTGAAAGCCCTTTTCCTAACTATGCTGACTCTGGCGCTGGTCAAGT		

DNA Sequence	CACAGGACACCGAAGAAACCATCACGTACACGCAATGCACTGACGGATATGAGTGGATCCTGTGAG ACAGCAATGCAAAGATATTGATGAATGTGACATTGTCCCAGACGCTTGTAAAGGTGGAATGAAGTGT GTCAACCACTATGGAGGATACCTCTGCCTTCCGAAAACAGCCCAGATTATTGTCAATAATGAACAGC CTCAGCAGGAAACACAACCAGCAGAAGGAACCTCAGGAGCAACCACCGGGGTTGTAGCTGCCAGCAG CATGGCAACCAGTGGAGTGTTGCCCGGGGGTGGTTTTGTGGCCAGTGCTGCTGCAGTCGCAGGCCCT GAAATGCAGACTGGCCGAAATAACTTTGTCATCCGGCGGAACCCAGCTGACCCTCAGCGCATTCCCT CCAACCCTTCCCACCGTATCCAGTGTGCAGCAGGCTACGAGCAAAGTGAACACAACGTGTGCCAAGA CATAGACGAGTGCACTGCAGGGACGCACAACTGTAGAGCAGACCAAGTGTGCATCAATTTACGGGGA TCCTTTGCATGTCAGTGCCCTCCTGGATATCAGAAGCGAGGGGAGCAGTGCGTAGATATAAATGAAT GTGATGCCAGCAATCAATGTGCTCAGCAGTGCTACAACATTCTTGGTTCATTCATCTGTCAGTGCAA TCAAGGATATGAGCTAAGCAGTGACAGGCTCAACTGTGAAGACATTGATGAATGCAGAACCTCAAGC TACCTGTGTCAATATCAATGTGTCAATGAACCTGGAAATTCTCATGTATGTGCCCCCAGGGATACC AAGTGGTGAGAAGTAGAACATGTCAAGATATAAATGAGTGTGAGACCACAAATGAATGCCGGGAGGA TGAAATGTGTTGGAATTATCATGGCGGCTTCCGTTGTTATCCACGAAATCCTTGTCAAGATCCCTAC ATTCTAACACCAGAGAACCGATGTGTTTGCCCAGTCTCAAATGCCATGTGCCGAGAACTGCCCCAGT CAATAGTCTACAAATACATGAGCATCCGATCTGATAGGTCTGTGCCATCAGACATCTTCCAGATACA GGCCACAACTATTTATGCCAACACCATCAATACTTTTCGGATTAAATCTGGAAATGAAAATGGAGAG TTCTACCTACGACAAACAAGTCCTGTAAGTGCAATGCTTGTGCTCGTGAAGTCATTATCAGGACCAA GAGAACATATCGTGGACCTGGAGATGCTGACAGTCAGCAGTATAGGGACCTTCCGCACAAGCTCTGT GTTAAGATTGACAATAATAGTGGGGCCATTTTCATTTTAGTCTTTTCTAAGAGTCAACCACAGGCAT TTAAGTCAGCCAAAGAATATTGTTACCTTAAAGCACTATTTTATTTATAGATATATCTAGTGCATCT ACATCTCTATACTGTACACTCACCCATAACAAACAATTACACCATGGTATAAAGTGGCATTTAATA TGTAAAGATTCAAAGTTTGTCTTTATTACTATATGTAAATTAGACATTAATCCACTAAACTGGTCTT CTTCAAGAGAGCTAAGTATACACTATCTGGTGAAACTTGGATTCTTTCCTATAAAAGTGGACCAAG CAATGATGATCTTCTGTGGTGCTTAAGGAAACTTACTAGAGCTCCACTAACAGTCTCATAAGGAGGC AGCCATCATAACCATTGAATAGCATGCAAGGGTAAGAATGAGTTTTTAACTGCTTTGTAAGAAAATG GAAAAGGTCAATAAAGATATATTTCTTTAGAAAATGGGGATCTGCCATATTTGTGTTGGTTTTTATT TTCATATCCAGCCTAAAGGTGGTTGTTTATTATATAGTAATAAATCATTGCTGTACAACATGCTGGT TTCTGTAGGGTATTTTTAATTTTGTCAGAAATTTTAGATTGTGAATATTTTGTAAAAAACAGTAAGC AAAATTTTCCAGAATTCCCAAAATGAACCAGATACCCCCTAGAAAATTATACTATTGAGAAATCTAT GGGGAGGATATGAGAAAATAAATTCCTTCTAAACCACATTGGAACTGACCTGAAGAAGCAAACTCGG AAAATATAATAACATCCCTGAATTCAGGCATTCACAAGATGCAGAACAAAATGGATAAAAGGTATTT CACTGGAGAAGTTTTAATTTCTAAGTAAAATTTAAATCCTAACACTTCACTAATTTATAACTAAAAT TTCTCATCTTCGTACTTGATGCTCACAGAGGAAGAAAATGATGATGGTTTTTATTCCTGGCATCCAG AGTGACAGTGAACTTAAGCAAATTACCCTCCTACCCAATTCTATGGAATATTTTATACGTCTCCTTG TTTAAAATCTGACTGCTTTACTTTGATGTATCATATTTTTAAATAAAAATAAATATTCCTTTAGAAG ATCACTCTAAAA		
	ORF Start: ATG at 153		ORF Stop: TAG at 1512

	SEQ ID NO: 32	453 aa	MW at 50198.0kD
NOV6b, CG121965-02 Protein Sequence	MLKALFLTMLTLALVKSQDTEETITYTQCTDGYEWDPVRQQCKDIDECDIVPDACKGGMKCVNHYGG YLCLPKTAQIIVNNEQPQQETQPAEGTSGATTGVVAASSMATSGVLPGGGFVASAAAVAGPEMQTGR NNFVIRRNPADPQRIPSNPSHRIQCAAGYEQSEHNVCQDIDECTAGTHNCRADQVCINLRGSFACQC PPGYQKRGEQCVDINECDASNQCAQQCYNILGSFICQCNQGYELSSDRLNCEDIDECRTSSYLCQYQ CVNEPGKFSCMCPQGYQVVRSRTCQDINECETTNECREDEMCWNYHGGFRCYPRNPCQDPYILTPEN RCVCPVSNAMCRELPQSIVYKYMSIRSDRSVPSDIFQIQATTIYANTINTFRIKSGNENGEFYLRQT SPVSAMLVLVKSLSGPREHIVDLEMLTVSSIGTFRTSSVLRLTIIVGPFSF		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 6B.

Table 6B. Comparison of NOV6a against NOV6b.

Protein Sequence	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV6b	1..493 1..453	440/493 (89%) 440/493 (89%)

Further analysis of the NOV6a protein yielded the following properties shown in Table 6C.

Table 6C. Protein Sequence Properties NOV6a	
PSort analysis:	0.3700 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 18 and 19

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6D.

Table 6D. Geneseq Results for NOV6a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV6a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB48077	Human extracellular signaling molecule (EXCS) (ID 1359783CD1) - Homo sapiens, 493 aa. [WO200070049-A2, 23-NOV-2000]	1..493 1..493	493/493 (100%) 493/493 (100%)	0.0
AAB72892	Human EFEMP1 - Homo sapiens, 493 aa. [WO200112823-A2, 22-FEB-2001]	1..493 1..493	493/493 (100%) 493/493 (100%)	0.0
AAG68188	Extracellular protein SEQ ID NO:104 - Homo sapiens, 387 aa. [WO200177327-A1, 18-OCT-2001]	107..493 1..387	387/387 (100%) 387/387 (100%)	0.0
AAY08066	Human EGF-like protein S1-5 fragment #1 encoded by GEN12205 cDNA - Homo sapiens, 350 aa. [WO9914241-A2, 25-MAR-1999]	144..493 1..350	350/350 (100%) 350/350 (100%)	0.0
AAY08490	Human EGF-like protein S1-5 fragment #2 encoded by GEN12205 cDNA - Homo sapiens, 348 aa. [WO9914241-A2, 25-MAR-1999]	3..346 1..348	344/348 (98%) 344/348 (98%)	0.0

In a BLAST search of public sequence datbases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6E.

Table 6E. Public BLASTP Results for NOV6a				
Protein Accession Number	**Protein/Organism/Length**	**NOV6a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**

Q12805	EGF-containing fibulin-like extracellular matrix protein 1 precursor (Fibulin-3) (FIBL-3) (Fibrillin-like protein) (Extracellular protein S1-5) - Homo sapiens (Human), 493 aa.	1..493 1..493	493/493 (100%) 493/493 (100%)	0.0
O35568	EGF-containing fibulin-like extracellular matrix protein 1 precursor (Fibulin-3) (FIBL-3) (T16 protein) - Rattus norvegicus (Rat), 493 aa.	1..493 1..493	459/493 (93%) 476/493 (96%)	0.0
I38449	extracellular protein - human, 387 aa.	107..493 1..387	387/387 (100%) 387/387 (100%)	0.0
AAH31184	Hypothetical protein - Mus musculus (Mouse), 387 aa.	107..493 1..387	371/387 (95%) 379/387 (97%)	0.0
Q9JM06	EGF-containing fibulin-like extracellular matrix protein 2 - Mus musculus (Mouse), 443 aa.	9..493 19..443	245/486 (50%) 311/486 (63%)	e-148

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table 6F.

Table 6F. Domain Analysis of NOV6a

Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
EGF	177..212	12/47 (26%) 29/47 (62%)	0.0002
EGF	218..252	14/47 (30%) 30/47 (64%)	0.0014
TIL	201..258	16/74 (22%) 34/74 (46%)	0.78
EGF	258..292	13/47 (28%) 23/47 (49%)	0.015

Example 7.

The NOV7 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis			
	SEQ ID NO: 33	1503 bp	
NOV7a, CG126129-01 DNA Sequence	GGACGCTGGATTAGAAGGCAGCAAAAAAAGATCTGTGCTGGCTGGAGCCCCCTCAGTGTGCAGGCTTA GAGGGACTAGGCTGGGTGTGGAGCTGCAGCGTATCCACAGGCCCCAGGATGCAGGCCCTGGTGCTACT CCTCTGCATTGGAGCCCTCCTCGGGCACAGCAGCTGGCAGAACCCTGCCAGCCCCCCGGAGGAGGGCT CCCCAGACCCCGACAGCACAGGGGCGCTGGTGGAGGAGGAGGATCCTTTCTTCAAAGTCGCCGTGAAC AAGCTGGCAGCGGCTGTCTCCAACTTCGGCTATGACCTGTACCGGGTGCGATCCAGCATGAGCCCCAC GACCAACGTGCTCCTGTCTCCTCTCAGTGTGGCCACGGCCCTCTCGGCCCTCTCGCTGGGAGCGGACG AGCGAACAGAATCCATCATTCACCGGGCTCTCTACTATGACTTGATCAGCAGCCCAGACATCCATGGT ACCTATAAGGAGCTCCTTGACACGGTCACTGCCCCCCAGAAGAACCTCAAGAGTGCCTCCCGGATCGT CTTTGAGAAGAAGCTGCGCATAAAATCCAGCTTTGTGGCACCTCTGGAAAAGTCATATGGGACCAGGC CCAGAGTCCTGACGGGCAACCCTCGCTTGGACCTGCAAGAGATCAACAACTGGGTGCAGGCGCAGATG AAAGGGAAGCTCGCCAGGTCCACAAAGGAAATTCCCGATGAGATCAGCATTCTCCTTCTCGGTGTGGC GCACTTCAAGGGGCAGTGGGTAACAAAGTTTGACTCCAGAAAGACTTCCCTCGAGGATTTCTACTTGG ATGAAGAGAGGACCGTGAGGGTCCCCATGATGTCGGACCCTAAGGCTGTTTTACGCTATGGCTTGGAT TCAGATCTCAGCTGCAAGATTGCCCAGCTGCCCTTGACCGGAAGCATGAGTATCATCTTCTTCCTGCC CCTGAAAGTGACCCAGAATTTGACCTTGATAGAGGAGAGCCTCACCTCCGAGTTCATTCATGACATAG ACCGAGAACTGAAGACCGTGCAGGCGGTCCTCACTGTCCCCAAGCTGAAGCTGAGTTACGAAGGCGAA GTCACCAAGTCCCTGCAGGAGATGAAGCTGCAATCCTTGTTTGATTCACCAGACTTTAGCAAGATCAC AGGCAAACCCATCAAGCTGACTCAGGTGGAACACCGGGCTGGCTTTGAGTGGAACGAGGATGGGGCGG GAACCACCCCCAGCCCAGGGCTGCAGCCTGCCCACCTCACCTTCCCGCTGGACTATCACCTTAACCAG CCTTTCATCTTCGTACTGAGGGACACAGACACAGGGGCCCTTCTCTTCATTGGCAAGATTCTGGACCC CAGGGGCCCCTAATATCCCAGTTTAATATTCCAATACCCTAGAAGAAAACCCGAGGACAGCAGATTC CACAGGACACGAAGGCTGCCCCTGTAAGGTTTCAATGCATACAATAAAAGAGCTTTATCCCTAAAAAA AAAAAAA		
	ORF Start: ATG at 117		ORF Stop: TAA at 1371

	SEQ ID NO: 34	418 aa	MW at 46385.6kD
NOV7a, CG126129-01 Protein Sequence	MQALVLLLCIGALLGHSSWQNPASPPEEGSPDPDSTGALVEEEDPFFKVAVNKLAAAVSNFGYDLYRV RSSMSPTTNVLLSPLSVATALSALSLGADERTESIIHRALYYDLISSPDIHGTYKELLDTVTAPQKNL KSASRIVFEKKLRIKSSFVAPLEKSYGTRPRVLTGNPRLDLQEINNWVQAQMKGKLARSTKEIPDEIS ILLLGVAHFKGQWVTKFDSRKTSLEDFYLDEERTVRVPMMSDPKAVLRYGLDSDLSCKIAQLPLTGSM SIIFFLPLKVTQNLTLIEESLTSEFIHDIDRELKTVQAVLTVPKLKLSYEGEVTKSLQEMKLQSLFDS PDFSKITGKPIKLTQVEHRAGFEWNEDGAGTTPSPGLQPAHLTFPLDYHLNQPFIFVLRDTDTGALLF IGKILDPRGP		

	SEQ ID NO: 35	368 bp	
NOV7b, CG126129-02 DNA Sequence	CTTAGAGGGACTAGGCTGGGTGTGGAGCTGCAGCGTATCCACAGGCCCCAGGATGCAGGCCCTGGTG CTACTCCTCTGCATTGGAGCCCTCCTCGGGCACAGCAGCTGCCAGAACCCTGCCAGCCCCCCGGAGG AGGGCTCCCCAGACCCCGACAGCACAGGGGCGCTGGTGGAGGAGGAGGATCCTTTCTTCAAAGTCCC CGTGAACAAGCTGGCAGCGGCTGTCTCCAACTTCGGCTATGACCTGTACCGGGTGCGATCCAGCGAA CAGAATCCATCATTCACCGGGCTCTCTACTATGACTTGATCAGCAGCCCAGACATCCATGGTACCTA TAAGGAGCTCCTTGACACGGTCACTGCCCCCCA		
	ORF Start: ATG at 53		ORF Stop: TGA at 305

	SEQ ID NO: 36	84 aa	MW at 8914.9kD
NOV7b, CG126129-02 Protein Sequence	MQALVLLLCIGALLGHSSCQNPASPPEEGSPDPDSTGALVEEEDPFFKVPVNKLAAAVSNFGYDLYR VRSSEQNPSFTGLSTMT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 7B.

Table 7B. Comparison of NOV7a against NOV7b.		
Protein Sequence	**NOV7a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV7b	16..71 16..71	40/56 (71%) 40/56 (71%)

Further analysis of the NOV7a protein yielded the following properties shown in Table 7C.

Table 7C. Protein Sequence Properties NOV7a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1443 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 16 and 17

A search of the NOV7a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 7D.

Table 7D. Geneseq Results for NOV7a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV7a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAR44800	Sequence of retinal pigmented epithelium-derived neurotrophic factor (PEDNF) - Homo sapiens, 418 aa. [WO9324529-A, 09-DEC-1993]	1..418 1..418	418/418 (100%) 418/418 (100%)	0.0
AAE10306	Human pigment epithelium derived growth factor (PEDF) - Homo sapiens, 418 aa. [WO200162725-A2, 30-AUG-2001]	1..418 1..418	416/418 (99%) 416/418 (99%)	0.0
AAR90287	Pigment epithelium-derived factor - Homo sapiens, 418 aa. [WO9533480-A1, 14-DEC-1995]	1..418 1..418	416/418 (99%) 416/418 (99%)	0.0

AAR90288	Modified pigment epithelium-derived factor (rPEDF) - Homo sapiens, 379 aa. [WO9533480-A1, 14-DEC-1995]	44..418 5..379	371/375 (98%) 374/375 (98%)	0.0
ABB57391	Rat mucocardial cell proliferation associated polypeptide SEQ ID NO 36 - Rattus norvegius, 418 aa. [WO200183705-A1, 08-NOV-2001]	1..415 1..415	343/416 (82%) 382/416 (91%)	0.0

In a BLAST search of public sequence datbases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7E.

Table 7E. Public BLASTP Results for NOV7a

Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
A47281	pigment epithelial-differentiating factor precursor - human, 418 aa.	1..418 1..418	416/418 (99%) 416/418 (99%)	0.0
P36955	Pigment epithelium-derived factor precursor (PEDF) (EPC-1) - Homo sapiens (Human), 418 aa.	1..418 1..418	414/418 (99%) 416/418 (99%)	0.0
Q96CT1	Hypothetical 46.4 kDa protein - Homo sapiens (Human), 418 aa.	1..418 1..418	413/418 (98%) 415/418 (98%)	0.0
O70629	Pigment epithelium-derived factor (Serine (Or cysteine) proteinase inhibitor, clade F (Alpha-2 antiplasmin, pigment epithelium derived factor). member 1) - Mus musculus (Mouse), 417 aa.	1..415 1..414	357/415 (86%) 391/415 (94%)	0.0
P97298	Pigment epithelium-derived factor precursor (PEDF) (Stromal cell- derived factor 3) (SDF-3) - Mus musculus (Mouse), 417 aa.	1..415 1..414	357/415 (86%) 391/415 (94%)	0.0

PFam analysis predicts that the NOV7a protein contains the domains shown in the Table 7F.

Table 7F. Domain Analysis of NOV7a			
Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
serpin	51..415	112/391 (29%) 262/391 (67%)	4.8e-83

Example 8.

The NOV8 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis			
	SEQ ID NO: 37	1154 bp	
NOV8a, CG142202-01 DNA Sequence	**ATG**GGGCGGCTGGTTCTGCTGTGGGGAGCTGCGGTCTTTCTGCTGGGAGGCTGGATGGCTTTGGGGCA AGGAGGAGCAGCAGAAGGAGTACAGATTCAGATCATCTACTTCAATTTAGAAACCGTGCAGGTGACAT GGAATGCCAGCAAATACTCCAGGACCAACCTGACTTTCCACTACAGATTCAACGGTGATGAGGCCTAT GACCAGTGCACCAACTACCTTCTCCAGGAAGGTCACACTTCGGGGTGCCTCCTAGACGCAGAGCAGCG AGACGACATTCTCTATTTCTCCATCAGGAATGGGACGCACCCCGTTTTCACCGCAAGTCGCTGGATGG TTTATTACCTGAAACCCAGTTCCCCGAAGCACGTGAGATTTTCGTGGCATCAGGATGCAGTGACGGTG ACGTGTTCTGACCTGTCCTACGGGGATCTCCTCTATGAGGTTCAGTACCGGAGCCCCTTCGACACCGA GTGGCAGTCCAAACAGGAAAATACCTGCAACGTCACCATAGAAGGCTTGGATGCCGAGAAGTGTTACT CTTTCTGGGTCAGGGTGAAGGCTATGGAGGATGTATATGGGCCAGACACATACCCAAGCGACTGGTCA GAGGTGACATGCTGGCAGAGAGGCGAGATTCGGGATGCCTGTGCAGAGACACCAACGCCTCCCAAACC AAAGCTGTCCAAATTTATTTTAATTTCCAGCCTGGCCATCCTTCTGATGGTGTCTCTCCTCCTTCTGT CTTTATGGAAATTATGGAGAGTGAGGAAGTTTCTCATTCCCAGCGTGCCAGACCCGAAATCCATCTTC CCCGGGCTCTTTGAGATACACCAAGGGAACTTCCAGGAGTGGATCACAGACACCCAGAACGTGGCCCA CCTCCACAAGATGGCAGGTGCAGAGCAAGAAAGTGGCCCCGAGGAGCCCCTGGTAGTCCAGTTGGCCA AGACTGAAGCCGAGTCTCCCAGGATGCTGGACCCACAGACCGAGGAGAAAGAGGCCTCTGGGGGATCC CTCCAGCTTCCCCACCAGCCCCTCCAAGGTGGTGATGTGGTCACAATCGGGGGCTTCACCTTTGTGAT GAATGACCGCTCCTACGTGGCGTTG**TGA**TGGACACACCACTGTCAAAGTCAACGTCAGAAGGGCGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1114

	SEQ ID NO: 38	371 aa	MW at 42040.3kD
NOV8a, CG142202-01 Protein Sequence	MGRLVLLWGAAVFLLGGWMALGQGGAAEGVQIQIIYFNLETVQVTWNASKYSRTNLTFHYRFNGDEAY DQCTNYLLQEGHTSGCLLDAEQRDDILYFSIRNGTHPVFTASRWMVYYLKPSSPKHVRFSWHQDAVTV TCSDLSYGDLLYEVQYRSPFDTEWQSKQENTCNVTIEGLDAEKCYSFWVRVKAMEDVYGPDTYPSDWS EVTCWQRGEIRDACAETPTPPKPKLSKFILISSLAILLMVSLLLLSLWKLWRVRKFLIPSVPDPKSIF PGLFEIHQGNFQEWITDTQNVAHLHKMAGAEQESGPEEPLVVQLAKTEAESPRMLDPQTEEKEASGGS LQLPHQPLQGGDVVTIGGFTFVMNDRSYVAL		

	SEQ ID NO: 39	1143 bp	
NOV8b, CG142202-03 DNA Sequence	**ATG**GGGCGGCTGGTTCTGCTGTGGGGAGCTGCGGTCTTTCTGCTGGGAGGCTGGATGGCTTTGGGGC AAGGAGGAGCAGAAGGAGTACAGATTCAGATCATCTACTTCAATTTAGAAACCGTGCAGGTGACATG GAATGCCAGCAAATACTCCAGGACCAACCTGACTTTCCACTACAGATTCAACGGTGATGAGGCCTAT GACCAGTGCACCAACTACCTTCTCCAGGAAGGTCACACTTCGGGGTGCCTCCTAGACGCAGAGCAGC GAGACGACATTCTCTATTTCTCCATCAGGAATGGGACGCACCCCGTTTTCACCGCAAGTCGCTGGAT GGTTTATTACCTGAAACCCAGTTCCCCGAAGCACGTGAGATTTTCGTGGCATCAGGATGCAGTGACG GTGACGTGTTCTGACCTGTCCTACGGGGATCTCCTCTATGAGGTTCAGTACCGGAGCCCCTTCGACA CCGAGTGGCAGTCCAAACAGGAAAATACCTGCAACGTCACCATAGAAGGCTTGGATGCCGAGAAGTG TTACTCTTTCTGGGTCAGGGTGAAGGCCATGGAGGATGTATATGGGCCAGACACATACCCAAGCGAC TGGTCAGAGGTGACATGCTGGCAGAGAGGCGAGATTCGGGATGCCTGTGCAGAGACACCAACGCCTC CCAAACCAAAGCTGTCCAAATTTATTTTAATTTCCAGCCTGGCCATCCTTCTGATGGTGTCTCTCCT CCTTCTGTCTTTATGGAAATTATGGAGAGTGAGGAAGTTTCTCATTCCCAGCGTGCCAGACCCGAAA TCCATCTTCCCCGGGCTCTTTGAGATACACCAAGGGAACTTCCAGGAGTGGATCACAGACACCCAGA		

	ACGTGGCCCACCTCCACAAGATGGCAGGTGCAGAGCAAGAAAGTGGCCCCGAGGAGCCCCTGGTAGT CCAGTTGGCCAAGACTGAAGCCGAGTCTCCCAGGATGCTGGACCCACAGACCGAGGAGAAAGAGGCC TCTGGGGGATCCCTCCAGCTTCCCCACCAGCCCCTCCAAGGTGGTGATGTGGTCACAATCGGGGGCT TCACCTTTGTGATGAATGACCGCTCCTACGTGGCGTTGTGATGGACACACCACTGTCAAAGTCAACG TCAG		
	ORF Start: ATG at 1		ORF Stop: TGA at 1111

	SEQ ID NO: 40	370 aa	MW at 41969.2kD
NOV8b, CG142202-03 Protein Sequence	MGRLVLLWGAAVFLLGGWMALGQGGAEGVQIQIIYFNLETVQVTWNASKYSRTNLTFHYRFNGDEAY DQCTNYLLQEGHTSGCLLDAEQRDDILYFSIRNGTHPVFTASRWMVYYLKPSSPKHVRFSWHQDAVT VTCSDLSYGDLLYEVQYRSPFDTEWQSKQENTCNVTIEGLDAEKCYSFWVRVKAMEDVYGPDTYPSD WSEVTCWQRGEIRDACAETPTPPKPKLSKFILISSLAILLMVSLLLLSLWKLWRVRKFLIPSVPDPK SIFPGLFEIHQGNFQEWITDTQNVAHLHKMAGAEQESGPEEPLVVQLAKTEAESPRMLDPQTEEKEA SGGSLQLPHQPLQGGDVVTIGGFTFVMNDRSYVAL		

	SEQ ID NO: 41	1154 bp	
NOV8c, CG142202-02 DNA Sequence	ATGGGGCGGCTGGTTCTGCTGTGGGGAGCTGCGGTCTTTCTGCTGGGAGGCTGGATGGCTTTGGGGCA AGGAGGAGCAGCAGAAGGAGTACAGATTCAGATCATCTACTTCAATTTAGAAACCGTGCAGGTGACAT GGAATGCCAGCAAATACTCCAGGACCAACCTGACTTTCCACTACAGATTCAACGGTGATGAGGCCTAT GACCAGTGCACCAACTACCTTCTCCAGGAAGGTCACACTTCGGGGTGCCTCCTAGACGCAGAGCAGCG AGACGACATTCTCTATTTCTCCATCAGGAATGGGACGCACCCCGTTTTCACCGCAAGTCGCTGGATGG TTTATTACCTGAAACCCAGTTCCCCGAAGCACGTGAGATTTTCGTGGCATCAGGATGCAGTGACGGTG ACGTGTTCTGACCTGTCCTACGGGGATCTCCTCTATGAGGTTCAGTACCGGAGCCCCTTCGACACCGA GTGGCAGTCCAAACAGGAAAATACCTGCAACGTCACCATAGAAGGCTTGGATGCCGAGAAGTGTTACT CTTTCTGGGTCAGGGTGAAGGCTATGGAGGATGTATATGGGCCAGACACATACCCAAGCGACTGGTCA GAGGTGACATGCTGGCAGAGAGGCGAGATTCGGGATGCCTGTGCAGAGACACCAACGCCTCCCAAACC AAAGCTGTCCAAATTTATTTTAATTTCCAGCCTGGCCATCCTTCTGATGGTGTCTCTCCTCCTTCTGT CTTTATGGAAATTATGGAGAGTGAGGAAGTTTCTCATTCCCAGCGTGCCAGACCCGAAATCCATCTTC CCCGGGCTCTTTGAGATACACCAAGGGAACTTCCAGGAGTGGATCACAGACACCCAGAACGTGGCCCA CCTCCACAAGATGGCAGGTGCAGAGCAAGAAAGTGGCCCCGAGGAGCCCCTGGTAGTCCAGTTGGCCA AGACTGAAGCCGAGTCTCCCAGGATGCTGGACCCACAGACCGAGGAGAAAGAGGCCTCTGGGGGATCC CTCCAGCTTCCCCACCAGCCCCTCCAAGGTGGTGATGTGGTCACAATCGGGGGCTTCACCTTTGTGAT GAATGACCGCTCCTACGTGGCGTTGTGATGGACACACCACTGTCAAAGTCAACGTCAGAAGGGCGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1114

	SEQ ID NO: 42	371 aa	MW at 42040.3kD
NOV8c, CG142202-02 Protein Sequence	MGRLVLLWGAAVFLLGGWMALGQGGAAEGVQIQIIYFNLETVQVTWNASKYSRTNLTFHYRFNGDEAY DQCTNYLLQEGHTSGCLLDAEQRDDILYFSIRNGTHPVFTASRWMVYYLKPSSPKHVRFSWHQDAVTV TCSDLSYGDLLYEVQYRSPFDTEWQSKQENTCNVTIEGLDAEKCYSFWVRVKAMEDVYGPDTYPSDWS EVTCWQRGEIRDACAETPTPPKPKLSKFILISSLAILLMVSLLLLSLWKLWRVRKFLIPSVPDPKSIF PGLFEIHQGNFQEWITDTQNVAHLHKMAGAEQESGPEEPLVVQLAKTEAESPRMLDPQTEEKEASGGS LQLPHQPLQGGDVVTIGGFTFVMNDRSYVAL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 8B.

Table 8B. Comparison of NOV8a against NOV8b and NOV8c.

Protein Sequence	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV8b	1..371 1..370	343/371 (92%) 343/371 (92%)
NOV8c	1..371 1..371	344/371 (92%) 344/371 (92%)

Further analysis of the NOV8a protein yielded the following properties shown in Table 8C.

Table 8C. Protein Sequence Properties NOV8a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.2473 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 23 and 24

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8D.

Table 8D. Geneseq Results for NOV8a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU77482	Human thymic stromal lymphopoietin receptor (TSLPR)-FLAG polypeptide - Homo sapiens, 379 aa. [WO200200724-A2, 03-JAN-2002]	1..371 1..371	370/371 (99%) 371/371 (99%)	0.0
AAU77481	Human TSLPR (thymic stromal lymphopoietin receptor) polypeptide - Homo sapiens, 371 aa. [WO200200724-A2, 03-JAN-2002]	1..371 1..371	370/371 (99%) 371/371 (99%)	0.0
AAU77220	Human thymic stromal lymphopoietin receptor(TSLPR)-FLAG protein sequence - Homo sapiens, 379 aa. [WO200200723-A2, 03-JAN-2002]	1..371 1..371	370/371 (99%) 371/371 (99%)	0.0
AAU77219	Human thymic stromal lymphopoietin receptor (TSLPR) protein sequence - Homo sapiens, 371 aa. [WO200200723-A2, 03-JAN-2002]	1..371 1..371	370/371 (99%) 371/371 (99%)	0.0

AAB71681	CRCGCL protein - Homo sapiens, 371 aa. [WO200112672-A2, 22-FEB-2001]	1..371 1..371	370/371 (99%) 371/371 (99%)	0.0

In a BLAST search of public sequence datbases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8E.

Table 8E. Public BLASTP Results for NOV8a

Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD26815	Sequence 7 from Patent WO0200723 - synthetic construct, 379 aa.	1..371 1..371	370/371 (99%) 371/371 (99%)	0.0
Q9HC73	Cytokine receptor CRL2 PRECUSOR (IL-XR) (Thymic stromal LYMPHOPOIETIN protein receptor TSLPR) - Homo sapiens (Human), 371 aa.	1..371 1..371	370/371 (99%) 371/371 (99%)	0.0
Q9H5R3	CDNA: FLJ23147 fis, clone LNG09295 - Homo sapiens (Human), 232 aa.	1..176 1..175	161/176 (91%) 166/176 (93%)	2e-93
Q8R4S8	Thymic stromal lymphopoietin receptor - Rattus norvegicus (Rat), 360 aa.	24..371 28..360	123/359 (34%) 183/359 (50%)	5e-48
Q9JMD5	Cytokine receptor delta1 - Mus musculus (Mouse), 359 aa.	6..371 1..359	135/380 (35%) 186/380 (48%)	4e-43

PFam analysis predicts that the NOV8a protein contains the domains shown in the Table 8F.

Table 8F. Domain Analysis of NOV8a

Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value
T-box	167..192	7/26 (27%) 22/26 (85%)	0.94

Example 9.

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis			
	SEQ ID NO: 43	828 bp	
NOV9a, CG142621-01 DNA Sequence	CTTATTAAAAACATACTCTTATTTTTCAGGATGTCAAACTTGGCACAATTTGACTCTGATTTTACCA ATCTAATTTTACTATTGATAACCAGGAGCAGAGTGGTAATGACTCTAATGCCTATGGAAATCTTTATG GATCTAGAAAGCAACAAGCTGGTGAGCAGCCTCAGCCTGCCTCCTTTGTTCCATCAGAGATGCTCATG TCATCGGGTTACGCAGGACAATTTTTTCAGCCAGCATCCAACTCAGATTATTATTCACAATCTCCTTA CATTGACAGTTTTGATGAAGAGCCTCCTTTGCTAGAAGAACTTGGAATCCATTTTGATCACATATGGC AAAAAACTTTGACAGTGTTAAACCCAATGAAGCCAGTAGATGGCAGCATTATGAATGAAACGGACCTC ACTGGACCCATTCTTTTTTGCGTAGCCCTGGGAGCCACCTTGCTTCTGGCAGGAAAAGTTCAGTTTGG TTATGTGTATGGCATGAGTGCCATTGGCTGCCTTGTGATTCATGCCTTGCTGAACCTGATGAGCTCTT CAGGGGTGTCGTACGGCTGTGTGGCCAGCGTGCTGGGTTACTGCCTGCTCCCCATGGTCATCCTGTCT GGTTGCGCCATGTTCTTTTCACTGCAGGGCATCTTTGGAATCATGTCATCCCTGGTCATCATTGGCTG GTGTAGTCTCTCAGCTTCCAAGATCTTCATTGCAGCCTTGCACATGGAAGGACAGCAGCTTCTTGTTG CCTACCCTTGTGCCATACTTTATGGACTTTTTGCCCTCCTAACAATTTTCTAAAGAATGTTTGAGATG GCATTTCAAGAC		
	ORF Start: ATG at 31		ORF Stop: TAA at 799

	SEQ ID NO: 44	256 aa	MW at 27775.6kD
NOV9a, CG142621-01 Protein Sequence	MSNLAQFDSDFYQSNFTIDNQEQSGNDSNAYGNLYGSRKQQAGEQPQPASFVPSEMLMSSGYAGQFFQ PASNSDYYSQSPYIDSFDEEPPLLEELGIHFDHIWQKTLTVLNPMKPVDGSIMNETDLTGPILFCVAL GATLLLAGKVQFGYVYGMSAIGCLVIHALLNLMSSSGVSYGCVASVLGYCLLPMVILSGCAMFFSLQG IFGIMSSLVIIGWCSLSASKIFIAALHMEGQQLLVAYPCAILYGLFALLTIF		

Further analysis of the NOV9a protein yielded the following properties shown in Table 9B.

Table 9B. Protein Sequence Properties NOV9a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9C.

Table 9C. Geneseq Results for NOV9a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB07505	Human GTP-binding protein (GTPB) (ID: 4879308CD1) - Homo sapiens, 257 aa. [WO200204510-A2, 17-JAN-2002]	1..256 1..257	160/259 (61%) 198/259 (75%)	2e-86
ABG34065	Human Pro peptide #36 - Homo sapiens, 257 aa. [WO200224888-A2, 28-MAR-2002]	1..256 1..257	160/259 (61%) 198/259 (75%)	2e-86
AAM41786	Human polypeptide SEQ ID NO 6717 - Homo sapiens, 260 aa. [WO200153312-A1, 26-JUL-2001]	1..256 4..260	160/259 (61%) 198/259 (75%)	2e-86
AAM40000	Human polypeptide SEQ ID NO 3145 - Homo sapiens, 257 aa. [WO200153312-A1, 26-JUL-2001]	1..256 1..257	160/259 (61%) 198/259 (75%)	2e-86
AAG67008	Human Yiplp28 polypeptide - Homo sapiens, 257 aa. [WO200166769-A1, 13-SEP-2001]	1..256 1..257	160/259 (61%) 198/259 (75%)	2e-86

In a BLAST search of public sequence datbases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9D.

Table 9D. Public BLASTP Results for NOV9a				
Protein Accession Number	**Protein/Organism/Length**	**NOV9a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9JIM5	YIP1B (2310016N21RIK protein) - Mus musculus (Mouse), 254 aa.	1..256 1..254	207/256 (80%) 225/256 (87%)	e-117
Q9EQQ2	Hypothetical 27.9 kDa protein (2610311I19Rik protein) (Similar to RIKEN cDNA 2310016N21 gene) - Mus musculus (Mouse), 257 aa.	1..256 1..257	160/259 (61%) 196/259 (74%)	3e-86

Q969M3	CDNA FLJ30014 fis, clone 3NB692000330, weakly similar to YIP1 protein (Similar to hypothetical protein AF140225) (Hypothetical 28.0 kDa protein) - Homo sapiens (Human), 257 aa.	1..256 1..257	160/259 (61%) 198/259 (75%)	5e-86
AAK67644	Golgi membrane protein SB140 - Homo sapiens (Human), 257 aa.	1..256 1..257	159/259 (61%) 197/259 (75%)	1e-84
Q9H338	Hypothetical 28.0 kDa protein - Homo sapiens (Human), 257 aa.	1..256 1..257	159/259 (61%) 195/259 (74%)	2e-84

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9E.

Table 9E. Domain Analysis of NOV9a			
Pfam Domain	**NOV9a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**

Example 10.

The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis			
	SEQ ID NO: 45	1837 bp	
NOV10a, CG142761-01 DNA Sequence	GGCACGAGGAACCCTTCCTGTTGCCTTAGGGGAACGTGGCTTTCCCTGCAGAGCCGGTGTCTCCGCC TGCGTCCCTGCTGCAGCAACCGGAGCTGGAGTCGGATCCCGAACGCACCCTCGCC**ATG**GACTCGGCC CTCAGCGATCCGCATAACGGCAGTGCCGAGGCAGGCGGCCCCACCAACAGCACTACGCGGCCGCCTT CCACGCCCGAGGGCATCGCGCTGGCCTACGGCAGCCTCCTGCTCATGGCGCTGCTGCCCATCTTCTT CGGCGCCCTGCGCTCCGTACGCTGCGCCCGCGGCAAGAATGCTTCAGACATGCCTGAAACAATCACC AGCCGGGATGCCGCCCGCTTCCCCATCATCGCCAGCTGCACACTCTTGGGGCTCTACCTCTTTTTCA AAATATTCTCCCAGGAGTACATCAACCTCCTGCTGTCCATGTATTTCTTCGTGCTGGGAATCCTGGC CCTGTCCCACACCATCAGCCCCTTCATGAATAAGTTTTTTCCAGCCAGCTTTCCAAATCGACAGTAC CAGCTGCTCTTCACACAGGGTTCTGGGGAAAACAAGGAAGAGATCATCAATTATGAATTTGACACCA AGGACCTGGTGTGCCTGGGCCTGAGCAGCATCGTTGGCGTCTGGTACCTGCTGAGGAAGCACTGGAT TGCCAACAACCTTTTTGGCCTGGCCTTCTCCCTTAATGGAGTAGAGCTCCTGCACCTCAACAATGTC AGCACTGGCTGCATCCTGCTGGGCGGACTCTTCATCTACGATGTCTTCTGGGTATTTGGCACCAATG TGATGGTGACAGTGGCCAAGTTCTTCGAGGCACCAATAAAATTGGTGTTTCCCCAGGATCTGCTGGA GAAAGGCCTCGAAGCAAACAACTTTGCCATGCTGGGACTTGGAGATGTCGTCATTCCAGGGATCTTC ATTGCCTTGCTGCTGCGCTTTGACATCAGCTTGAAGAAGAATACCCACACCTACTTCTACACCAGCT TTGCAGCCTACATCTTCGGCCTGGGCCTTACCATCTTCATCATGCACATCTTCAAGCATGCTCAGCC TGCCCTCCTATACCTGGTCCCCGCCTGCATCGGTTTTCCTGTCCTGGTGGCGCTGGCCAAGGGAGAA GTGACAGAGATGTTCAGCTACGAGTCCTCGGCGGAAATCCTGCCTCATACCCCGAGGCTCACCCACT TCCCCACAGTCTCGGGCTCCCCAGCCAGCCTGGCCGACTCCATGCAGCAGAAGCTAGCTGGCCCTCG		

	CCGCCGGCGCCCGCAGAATCCCAGCGCCATGTAATGCCCAGCGGGTGCCCACCTGCCCGCTTCCCCC TACTGCCCCGGGGCCCAAGTTATGAGGAGTCAAATCCTAAGGATCCAGCGGCAGTGACAGAATCCAA AGAGGGAACAGAGGCATCAGCATCGAAGGGGCTGGAGAAGAAAGAGAAATGATGCAGCTGGTGCCCG AGCCTCTCAGGGCCAGACCAGACAGATGGGGGCTGGGCCCACACAGGCGTGCACCGGTAGAGGGCAC AGGAGGCCAAGGGCAGCTCCAGGACAGGGCAGGGGGCAGCAGGATACCTCCAGCCAGGCCTCTGTGG CCTCTGTTTCCTTCTCCCTTTCTTGGCCCTCCTCTGCTCCTCCCCACACCCTGCAGGCAAAAGAAAC CCCCAGCTTCCCCCCTCCCCGGGAGCCAGGTGGGAAAAGTGGGTGTGATTTTTAGATTTTGTATTGT GGACTGATTTTGCCTCACATTAAAAACTCATCCCATGGCCAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 123		ORF Stop: TAA at 1305

	SEQ ID NO: 46	394 aa	MW at 43482.2kD
NOV10a, CG142761-01 Protein Sequence	MDSALSDPHNGSAEAGGPTNSTTRPPSTPEGIALAYGSLLLMALLPIFFGALRSVRCARGKNASDMP ETITSRDAARFPIIASCTLLGLYLFFKIFSQEYINLLLSMYFFVLGILALSHTISPFMNKFFPASFP NRQYQLLFTQGSGENKEEIINYEFDTKDLVCLGLSSIVGVWYLLRKHWIANNLFGLAFSLNGVELLH LNNVSTGCILLGGLFIYDVFWVFGTNVMVTVAKFFEAPIKLVFPQDLLEKGLEANNFAMLGLGDVVI PGIFIALLLRFDISLKKNTHTYFYTSFAAYIFGLGLTIFIMHIFKHAQPALLYLVPACIGFPVLVAL AKGEVTEMFSYESSAEILPHTPRLTHFPTVSGSPASLADSMQQKLAGPRRRRPQNPSAM		

Further analysis of the NOV10a protein yielded the following properties shown in Table 10B.

Table 10B. Protein Sequence Properties NOV10a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 61 and 62

A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 10C.

Table 10C. Geneseq Results for NOV10a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV10a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB88567	Human hydrophobic domain containing protein clone HP03010 #31 - Homo sapiens, 377 aa. [WO200112660-A2, 22-FEB-2001]	1..379 1..375	353/379 (93%) 359/379 (94%)	0.0
AAB10549	Human aspartate protease psl 3 protein - Homo sapiens, 377 aa. [WO200043505-A2, 27-JUL-2000]	1..379 1..375	353/379 (93%) 359/379 (94%)	0.0

AAY27132	Human glioblastoma-derived polypeptide (clone OA004FG) - Homo sapiens, 377 aa. [WO9933873-A1, 08-JUL-1999]	1..379 1..375	353/379 (93%) 359/379 (94%)	0.0
AAM93670	Human polypeptide, SEQ ID NO: 3554 - Homo sapiens, 377 aa. [EP1130094-A2, 05-SEP-2001]	1..379 1..375	352/379 (92%) 359/379 (93%)	0.0
AAY27133	Human glioblastoma-derived polypeptide (clone OA004LD) - Homo sapiens, 377 aa. [WO9933873-A1, 08-JUL-1999]	1..379 1..375	351/379 (92%) 357/379 (93%)	0.0

In a BLAST search of public sequence datbases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10D.

Table 10D. Public BLASTP Results for NOV10a				
Protein Accession Number	**Protein/Organism/Length**	**NOV10a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q95H87	Similar to histocompatibility 13 - Homo sapiens (Human), 377 aa.	1..379 1..375	354/379 (93%) 360/379 (94%)	0.0
Q8TCT9	Signal peptide peptidase - Homo sapiens (Human), 377 aa.	1..379 1..375	353/379 (93%) 359/379 (94%)	0.0
BAC11519	CDNA FLJ90802 fis, clone Y79AA1000226 - Homo sapiens (Human), 377 aa.	1..379 1..375	352/379 (92%) 359/379 (93%)	0.0
Q9D8V0	1200006O09Rik protein - Mus musculus (Mouse), 378 aa.	1..349 1..349	335/349 (95%) 343/349 (97%)	0.0
AAM22075	Minor histocompatibility antigen H13 isoform 1 - Mus musculus (Mouse), 378 aa.	1..379 1..376	339/379 (89%) 352/379 (92%)	0.0

PFam analysis predicts that the NOV10a protein contains the domains shown in the Table 10E.

Table 10E. Domain Analysis of NOV10a

Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 11.

The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis			
	SEQ ID NO: 47	615 bp	
NOV11a, CG143926-01 DNA Sequence	CTCAGAGTCTCCTCAGACGCCGAGATGCGGGTCACGGCACCCCGAACCGTCCTCCTGCTGCTCTCGG CGGCCCTGGCCCTGACCGAGTGCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGACAAGCT GGAGCGCGCTGACCCCCCAAAGACACACGTGACCCACCACCCCATCTCTGACCATGAGGCCACCCTG AGGTGCTGGGCCCTGGGTTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGCGAGGACC AAACTCAGGACACTGAGCTTGTGGAGACCAGACCAGCAGGAGATAGAACCTTCCAGAAGTGGGCAGC TGTGGTGGTGCCTTCTGGAGAAGAGCAGAGATACACATGCCATGTACAGCATGAGGGGCTGCCGAAG CCCCTCACCCTGAGATGGGAGCCGTCTTCCCAGTCCACCGTCCCCATCGTGGGCATTGTTGCTGGCC TGGCTGTCCTAGCAGTTGTGGTCATCGGAGCTGTGGTCGCTGCTGTGATGTGTAGGAGGAAGAGTTC AGGTGGAAAAGGAGGGAGCTACTCTCAGGCTGCGTGCAGCGACAGTGCCCAGGGCTCTGATGTGTCT CTCACAGCTTGA		
	ORF Start: ATG at 25		ORF Stop: TGA at 613

	SEQ ID NO: 48	196 aa	MW at 21301.0kD
NOV11a, CG143926-01 Protein Sequence	MRVTAPRTVLLLLSAALALTECVEWLRRYLENGKDKLERADPPKTHVTHHPISDHEATLRCWALGFY PAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEP SSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDVSLTA		

Further analysis of the NOV11a protein yielded the following properties shown in Table 11B.

Table 11B. Protein Sequence Properties NOV11a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 23 and 24

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11C.

Table 11C. Geneseq Results for NOV11a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAP70155	Sequence encoded by genomic DNA encoding human histocompatibility antigen HLA-B 27 - Homo sapiens, 362 aa. [EP226069-A, 24-JUN-1987]	21..196 187..362	173/176 (98%) 175/176 (99%)	e-100
AAP70590	Sequence of the human histocompatibility antigen HLA B27 - Homo sapiens, 337 aa. [DE3542024-A, 04-JUN-1987]	21..196 162..337	172/176 (97%) 174/176 (98%)	e-99
AAR03144	Sequence of HLA-B51 antigen - Homo sapiens, 362 aa. [EP354580-A, 14-FEB-1990]	22..196 188..362	167/175 (95%) 172/175 (97%)	4e-97
AAR03142	Sequence of HLA-Bw52 antigen - Homo sapiens, 362 aa. [EP354580-A, 14-FEB-1990]	22..196 188..362	167/175 (95%) 172/175 (97%)	4e-97
AAU32882	Novel human secreted protein #3373 - Homo sapiens, 369 aa. [WO200179449-A2, 25-OCT-2001]	22..196 191..366	169/176 (96%) 170/176 (96%)	1e-95

In a BLAST search of public sequence datbases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11D.

Table 11D. Public BLASTP Results for NOV11a				
Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q31603	Lymphocyte antigen - Homo sapiens (Human), 362 aa.	21..196 187..362	176/176 (100%) 176/176 (100%)	e-101
Q29854	HLA-B alpha chain antigen precursor - Homo sapiens (Human), 362 aa.	21..196 187..362	176/176 (100%) 176/176 (100%)	e-101
Q29861	HLA-BPOT (classI) - Homo sapiens (Human), 362 aa.	21..196 187..362	176/176 (100%) 176/176 (100%)	e-101

Q29681	MHC class I antigen heavy chain precursor - Homo sapiens (Human), 362 aa.	21..196 187..362	176/176 (100%) 176/176 (100%)	e-101
Q29638	MHC class I antigen - Homo sapiens (Human), 362 aa.	21..196 187..362	176/176 (100%) 176/176 (100%)	e-101

PFam analysis predicts that the NOV11a protein contains the domains shown in the Table 11E.

Table 11E. Domain Analysis of NOV11a

Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
MHC_I	20..37	15/18 (83%) 17/18 (94%)	1.5e-07
ig	54..119	15/67 (22%) 48/67 (72%)	2.8e-09

Example 12.

The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis

	SEQ ID NO: 49	555 bp	
NOV12a, CG144193-01 DNA Sequence	ATGATTTCCAGAATGGAGAAGATGACGATGATGATGAAGATATTGATTATGTTTGCTCTTGGAATGA ACTACTGGTCTTGCTCAGGTTTCCCAGTGTACGACTACGATCCATCCTCCTTAAGGGATGCCCTCAG TGCCTCTGTGGTAAAAGTGAATTCCCAGTCACTGAGTCCGTATCTGTTTCGGGCATTCAGAAGCTCA TTAAAAAGAGTTGAGGTCCTAGATGAGAACAACTTGGTCATGAATTTAGAGTTCAGCATCCGGGAGA CAACATGCAGGAAGGATTCTGGAGAAGATCCCGCTACATGTGCCTTCCAGAGGGACTACTATGTGTC CACGTCTGAGTCTTACAGCAGCGAAGAGATGATTTTTGGGGACATGTTGGGATCTCATAAATGGAGA AGCAATTATCTATTTGGTCTCATTTCAGACGAGTCCATAAGTGAACAATTTTATGATCGGTCACTTG GGATCATGAGAAGGGTATTGCCTCCTGGAAACAGAAGGTACCCAAACCACCGGCACAGAGCAAGAAT AAATACTGACTTTGAGTAA		
	ORF Start: ATG at 1		ORF Stop: TAA at 553

	SEQ ID NO: 50	184 aa	MW at 21465.1kD
NOV12a, CG144193-01 Protein Sequence	MISRMEKMTMMMKILIMFALGMNYWSCSGFPVYDYDPSSLRDALSASVVKVNSQSLSPYLFRAFRSS LKRVEVLDENNLVMNLEFSIRETTCRKDSGEDPATCAFQRDYYVSTSESYSSEEMIFGDMLGSHKWR SNYLFGLISDESISEQFYDRSLGIMRRVLPPGNRRYPNHRHRARINTDFE		

	SEQ ID NO: 51	636 bp	
NOV12b,	ATGATTTCCAGAATGGAGAAGATGACGATGATGATGAAGATATTGATTATGTTTGCTCTTGGAATGA ACTACTGGTCTTGCTCAGGTTTCCCAGTGTACGACTACGATCCATCCTCCTTAAGGGATGCCCTCAG		

CG144193-02 DNA Sequence	TGCCTCTGTGGTAAAAGTGAATTCCCAGTCACTGAGTCCGTATCTGTTTCGGGCATTCAGAAGCTCA TTAAAAAGAGTTGAGGTCCTAGATGAGAACAACTTGGTCATGAATTTAGAGTTCAGCATCCGGGAGA CAACATGCAGGAAGGATTCTGGAGAAGATCCCGCTACATGTGCCTTCCAGAGGGACTACTATGTGTC CACAGCTGTTTGCAGAAGCACCGTGAAGGTATCTGCCCAGCAGGTGCAGGGCGTGCATGCTCGCTGC AGCTGGTCCTCCTCCACGTCTGAGTCTTACAGCAGCGAAGAGATGATTTTTGGGGACATGTTGGGAT CTCATAAATGGAGAAACAATTATCTATTTGGTCTCATTTCAGACGAGTCCATAAGTGAACAATTTTA TGATCGGTCACTTGGGATCATGAGAAGGGTATTGCCTCCTGGAAACAGAAGGTACCCAAACCACCGG CACAGAGCAAGAATAAATACTGACTTTGAGTAA		
	ORF Start: ATG at 1		ORF Stop: TAA at 634

	SEQ ID NO: 52	211 aa	MW at 24337.4kD
NOV12b, CG144193-02 Protein Sequence	MISRMEKMTMMMKILIMFALGMNYWSCSGFPVYDYDPSSLRDALSASVVKVNSQSLSPYLFRAFRSS LKRVEVLDENNLVMNLEFSIRETTCRKDSGEDPATCAFQRDYYVSTAVCRSTVKVSAQQVQGVHARC SWSSSTSESYSSEEMIFGDMLGSHKWRNNYLFGLISDESISEQFYDRSLGIMRRVLPPGNRRYPNHR HRARINTDFE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 12B.

Table 12B. Comparison of NOV12a against NOV12b.		
Protein Sequence	**NOV12a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV12b	1..184 1..211	176/211 (83%) 180/211 (84%)

Further analysis of the NOV12a protein yielded the following properties shown in Table 12C.

Table 12C. Protein Sequence Properties NOV12a	
PSort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 30 and 31

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12D.

Table 12D. Geneseq Results for NOV12a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV12a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**

AAR10321	Human BMP - Homo sapiens, 211 aa. [EP409472-A, 23-JAN-1991]	1..184 1..211	183/211 (86%) 184/211 (86%)	e-100
AAR10320	Human BMP - Homo sapiens, 211 aa. [EP409472-A, 23-JAN-1991]	1..184 1..211	183/211 (86%) 184/211 (86%)	e-100
AAR10319	Bovine BMP - Bos taurus, 203 aa. [EP409472-A, 23-JAN-1991]	5..184 1..203	117/206 (56%) 140/206 (67%)	2e-55
AAW02632	Bovine phosphoprotein Spp24 - Bos taurus, 200 aa. [WO9621006-A1, 11-JUL-1996]	10..184 1..200	113/201 (56%) 137/201 (67%)	1e-54
AAR10317	Bovine BMP - exon 3 - Bos taurus, 41 aa. [EP409472-A, 23-JAN-1991]	71..111 1..41	26/41 (63%) 33/41 (80%)	2e-08

In a BLAST search of public sequence datbases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12E.

Table 12E. Public BLASTP Results for NOV12a

Protein Accession Number	Protein/Organism/Length	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q13103	Secreted phosphoprotein 24 precursor (SPP-24) - Homo sapiens (Human), 211 aa.	1..184 1..211	183/211 (86%) 184/211 (86%)	3e-99
AAH27494	RIKEN cDNA 0610038O04 gene - Mus musculus (Mouse), 203 aa.	11..184 5..203	121/200 (60%) 143/200 (71%)	4e-59
Q9DCG1	0610038O04Rik protein - Mus musculus (Mouse), 203 aa.	11..184 5..203	121/200 (60%) 143/200 (71%)	4e-59
Q27967	Secreted phosphoprotein 24 precursor (SPP-24) - Bos taurus (Bovine), 200 aa.	10..184 1..200	114/201 (56%) 137/201 (67%)	2e-54
Q62740	Secreted phosphoprotein 24 (SPP-24) - Rattus norvegicus (Rat), 180 aa.	30..184 1..180	109/181 (60%) 128/181 (70%)	7e-51

PFam analysis predicts that the NOV12a protein contains the domains shown in the Table 12F.

Table 12F. Domain Analysis of NOV12a			
Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Cathelicidins	37..104	18/69 (26%) 33/69 (48%)	0.15
cystatin	30..106	17/83 (20%) 51/83 (61%)	0.14

Example 13.

The NOV13 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis			
	SEQ ID NO: 53	835 bp	
NOV13a, CG144545-01 DNA Sequence	CCTCTCTCTCTGACTGCCTGCTGGAA**ATG**CCCCCATCTCCCTTTGAGTCCTCCTCCCGGGCGACTCC TGTGACCTGTAACCTCTGTCCTGAAATCATCACAATGGCCAGGGTGGCCTCAGCTCAGGGCCTCTGT GACATCACCAAGGGCCTGGCACCAGGTGCCCAGTCTCCCAGTTGCGAGGGCAAGCAAACCCGTCATG AGCAACTCCCTTCCCCATCTCTGCTCACCATGTGGACGCTGAAATCGTCCCTGGTCCTGCTTCTGTG CCTCACCTGCAGCTATGCCTTTATGTTCTCTTCTCTGAGACAGAAAACTAGCGAACCCCAGGGGAAG GTGCAATACGGAGAGCACTTTCGGATTCGGCAGAATCTACCAGAGCACACCCAAGGCTGGCTTGGGA GCAAATGGCTCTGGCTTCTTTTTGTTGTTGTGCCGTTTGTGATACTGCAGTGTCAAAGAGACAGTGA GAAGAATAAGGAGCAGAGTCCTCCTGGCCTTCGAGGCGGCCAACTTCACTCTCCATTAAAGAAAAAA AGAAATGCTTCCCCCAACAAAGACTGTGCATTCAATACCTTAATGGAACTCGAGGTGGAGCTTATGA AATTTGTGTCCGAAGTGCGGAATCTTAAAGGTGCCATGGCAACAGGTAGTGGCAGTAACCTCAGGCT TCGAAGGTCAGAGATGCCTGCAGATCCATACCATGTCACGATCTGTGAAATATGGGGAGAAGAAAGC TCTAGC**TGA**ATGGATTTGTGTGTCAGGAGAGAAAAAGTTGAGTGTTGACAAACTGTATGCAAACTA ATAAAACTATTCTGAAGAAAAGAAAAAAAAA		
	ORF Start: ATG at 27		ORF Stop: TGA at 744

	SEQ ID NO: 54	239 aa	MW at 26610.4kD
NOV13a, CG144545-01 Protein Sequence	MPPSPFESSSRATPVTCNLCPEIITMARVASAQGLCDITKGLAPGAQSPSCEGKQTRHEQLPSPSLL TMWTLKSSLVLLLCLTCSYAFMFSSLRQKTSEPQGKVQYGEHFRIRQNLPEHTQGWLGSKWLWLLFV VVPFVILQCQRDSEKNKEQSPPGLRGGQLHSPLKKKRNASPNKDCAFNTLMELEVELMKFVSEVRNL KGAMATGSGSNLRLRRSEMPADPYHVTICEIWGEESSS		

Further analysis of the NOV13a protein yielded the following properties shown in Table 13B.

Table 13B. Protein Sequence Properties NOV13a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13C.

Table 13C. Geneseq Results for NOV13a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU68550	Human novel cytokine encoded by cDNA 790CIP2D_11 #1 - Homo sapiens, 239 aa. [WO200175093-A1, 11-OCT-2001]	1..239 1..239	235/239 (98%) 237/239 (98%)	e-137
AAY53032	Human secreted protein clone di393_2 protein sequence SEQ ID NO:70 - Homo sapiens, 171 aa. [WO9957132-A1, 11-NOV-1999]	69..239 1..171	168/171 (98%) 170/171 (99%)	1e-96
AAG00463	Human secreted protein, SEQ ID NO: 4544 - Homo sapiens, 101 aa. [EP1033401-A2, 06-SEP-2000]	69..169 1..101	100/101 (99%) 100/101 (99%)	5e-55
AAY12683	Human 5' EST secreted protein SEQ ID NO:273 - Homo sapiens, 101 aa. [WO9906549-A2, 11-FEB-1999]	69..169 1..101	100/101 (99%) 100/101 (99%)	5e-55
AAM87953	Human immune/haematopoietic antigen SEQ ID NO:15546 - Homo sapiens, 89 aa. [WO200157182-A2, 09-AUG-2001]	151..239 1..89	85/89 (95%) 88/89 (98%)	4e-44

In a BLAST search of public sequence datbases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13D.

Table 13D. Public BLASTP Results for NOV13a

Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9HCV6	DJ1153D9.4 (Novel protein) - Homo sapiens (Human), 138 aa (fragment).	102..239 1..138	120/138 (86%) 126/138 (90%)	3e-66
Q9D9T2	1700029J11Rik protein - Mus musculus (Mouse), 170 aa.	72..238 5..169	101/168 (60%) 122/168 (72%)	2e-46
Q9HCV7	DJ1153D9.3 (novel protein) - Homo sapiens (Human), 94 aa.	69..154 1..86	84/86 (97%) 84/86 (97%)	4e-44
Q96C09	Similar to neuronal thread protein - Homo sapiens (Human), 106 aa.	69..156 1..88	80/88 (90%) 82/88 (92%)	8e-42
Q8YR98	Hypothetical protein All3550 - Anabaena sp. (strain PCC 7120), 208 aa.	9..61 21..71	18/53 (33%) 31/53 (57%)	2.6

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13E.

Table 13E. Domain Analysis of NOV13a			
Pfam Domain	NOV13a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 14.

The NOV14 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 55	855 bp	
NOV14a, CG144884-01 DNA Sequence	CTGCGTTGCTGGGAAGTTCTGGAAGGAAGCATGTGCTCCAGAGGTTGGGATTCGTGTCTGGCTCTGG AATTGCTACTGCTGCCTCTGTCACTCCTGGTGACCAGCATTCAAGGTCACTTGGTACATATGACCGT GGTCTCCGGCAGCAACGTGACTCTGAACATCTCTGAGAGCCTGCCTGAGAACTACAAACAACTAACC TGGTTTTATACTTTCGACCAGAAGATTGTAGAATGGGATTCCAGAAAATCTAAGTACTTTGAATCCA AATTTAAAGGCAGGGTCAGACTTGATCCTCAGAGTGGCGCACTGTACATCTCTAAGGTCCAGAAAGA GGACAACAGCACCTACATCATGAGGGTGTTGAAAAAGACTGGGAATGAGCAAGAATGGAAGATCAAG CTGCAAGTGCTTGACCCTGTACCCAAGCCTGTCATCAAAATTGAGAAGATAGAAGACATGGATGACA ACTGTTATCTGAAACTGTCATGTGTGATACCTGGCGAGTCTGTAAACTACACCTGGTATGGGGACAA AAGGCCCTTCCCAAAGGAGCTCCAGAACAGTGTGCTTGAAACCACCCTTATGCCACATAATTACTCC AGGTGTTATACTTGCCAAGTCAGCAATTCTGTGAGCAGCAAGAATGGCACGGTCTGCCTCAGTCCAC CCTGTACCCTGGCCCGGTCCTTTGGAGTAGAATGGATTGCAAGTTGGCTAGTGGTCACGGTGCCCAC		

	CATTCTTGGCCTGTTACTTACCTGAGATGAGCTCTTTTAACTCAAGCGAAACTTCAAGGCCAGAAGA TCTTGCCTGTTGGTGATCATGCTCCTCACCAGGACAGAGACTGTATAAAGG		
	ORF Start: ATG at 31		ORF Stop: TGA at 760

	SEQ ID NO: 56	243 aa	MW at 27682.8kD
NOV14a, CG144884-01 Protein Sequence	MCSRGWDSCLALELLLLPLSLLVTSIQGHLVHMTVVSGSNVTLNISESLPENYKQLTWFYTFDQKIV EWDSRKSKYFESKFKGRVRLDPQSGALYISKVQKEDNSTYIMRVLKKTGNEQEWKIKLQVLDPVPKP VIKIEKIEDMDDNCYLKLSCVIPGESVNYTWYGDKRPFPKELQNSVLETTLMPHNYSRCYTCQVSNS VSSKNGTVCLSPPCTLARSFGVEWIASWLVVTVPTILGLLLT		

	SEQ ID NO: 57	573 bp	
NOV14b, CG144884-02 DNA Sequence	GGAAGTTCTGGAAGCAAGCATGTGCTCCAGAGGTTGGGATTCGTGTCTGGCTCTGGAATTGCTACTG CTGCCTCTGTCACTCCTGGTGACCAGCATTCAAGGTCACTTGGTACATATGACCGTGGTCTCCGGCA GCAACGTGACTCTGAACATCTCTGAGAGCCTGCCTGAGAACTACAAACAACTAACCTGGTTTTATAC TTTCGACCAGAAGATTGTAGAATGGGATTCCAGAAAATCTAAGTACTTTGAATCCAAATTTAAAGGC AGGGTCAGACTTGATCCTCAGAGTGGCGCACTGTACATCTCTAAGGTCCAGAAAGAGGACAACAGCA CCTACATCATGAGGGTGTTGAAAAAGACTGGGAATGAGCAAGAATGGAAGATCAAGCTGCAAGTGCT TGCCCGGTCCTTTGGAGTAGAATGGATTGCAAGTTGGCTAGTGGTCACGGTGCCCACCATTCTTGGC CTGTTACTTACCTGAGATGAGCTCTTTTAACTCAAGCGAAACTTCAAGGCCAGAAGATCTTGCCTGT TGGTGATCATGCTCCTCACCAGGACAGAGACTGTATA		
	ORF Start: ATG at 20		ORF Stop: TGA at 482

	SEQ ID NO: 58	154 aa	MW at 17670.4kD
NOV14b, CG144884-02 Protein Sequence	MCSRGWDSCLALELLLLPLSLLVTSIQGHLVHMTVVSGSNVTLNISESLPENYKQLTWFYTFDQKIV EWDSRKSKYFESKFKGRVRLDPQSGALYISKVQKEDNSTYIMRVLKKTGNEQEWKIKLQVLARSFGV EWIASWLVVTVPTILGLLLT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 14B.

Table 14B. Comparison of NOV14a against NOV14b.		
Protein Sequence	**NOV14a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV14b	1..128 1..128	115/128 (89%) 115/128 (89%)

Further analysis of the NOV14a protein yielded the following properties shown in Table 14C.

Table 14C. Protein Sequence Properties NOV14a	
PSort analysis:	0.9190 probability located in plasma membrane; 0.2000 probability located in lysosome (membrane); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 29 and 30

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14D.

Table 14D. Geneseq Results for NOV14a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV14a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU74426	Human protein sequence #4, related to isolation of genes within SLE-1B - Homo sapiens, 243 aa. [WO200188200-A2, 22-NOV-2001]	1..243 1..243	242/243 (99%) 242/243 (99%)	e-141
AAW35857	Human CD48 for use in T lymphocyte veto molecule - Homo sapiens, 194 aa. [WO9737687-A1, 16-OCT-1997]	27..220 1..194	194/194 (100%) 194/194 (100%)	e-113
AAU74427	Mouse protein sequence #4, related to isolation of genes within SLE-1B - Mus musculus, 240 aa. [WO200188200-A2, 22-NOV-2001]	1..243 1..240	129/247 (52%) 163/247 (65%)	2e-60
AAG00342	Human secreted protein, SEQ ID NO: 4423 - Homo sapiens, 111 aa. [EP1033401-A2, 06-SEP-2000]	1..111 1..111	109/111 (98%) 109/111 (98%)	4e-58
ABG47129	Human peptide encoded by genome-derived single exon probe SEQ ID 36794 - Homo sapiens, 96 aa. [WO200186003-A2, 15-NOV-2001]	33..128 1..96	96/96 (100%) 96/96 (100%)	4e-50

In a BLAST search of public sequence datbases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14E.

Table 14E. Public BLASTP Results for NOV14a

Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P09326	B-lymphocyte activation marker BLAST-1 precursor (BCM1 surface antigen) (Leucocyte antigen MEM-102) (TCT.1) (Antigen CD48) - Homo sapiens (Human), 243 aa.	1..243 1..243	243/243 (100%) 243/243 (100%)	e-142
AAH30224	Similar to B-lymphocyte activation marker BLAST-1 (BCM1 surface antigen) (Leucocyte antigen MEM-102) (TCT.1) (Antigen CD48) - Homo sapiens (Human), 169 aa.	1..148 1..148	132/148 (89%) 134/148 (90%)	1e-69
P18181	MRC OX-45 surface antigen precursor (BCM1 surface antigen) (BLAST-1) (CD48) (HM48-1) - Mus musculus (Mouse), 240 aa.	1..243 1..240	129/247 (52%) 163/247 (65%)	5e-60
P10252	MRC OX-45 surface antigen precursor (BCM1 surface antigen) (BLAST-1) (CD48) - Rattus norvegicus (Rat), 240 aa.	10..242 10..239	120/235 (51%) 155/235 (65%)	2e-56
Q8VE93	Similar to RIKEN cDNA 2310026I04 gene - Mus musculus (Mouse), 285 aa.	42..213 35..221	51/187 (27%) 85/187 (45%)	1e-09

PFam analysis predicts that the NOV14a protein contains the domains shown in the Table 14F.

Table 14F. Domain Analysis of NOV14a

Pfam Domain	NOV14a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ig	147..198	10/56 (18%) 36/56 (64%)	0.011

Example 15.

The NOV15 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis			
	SEQ ID NO: 59	700 bp	
NOV15a, CG145122-01 DNA Sequence	GGGATCCGACTCTAGTCGTAATGGAGGCGGGCGGCTTTCTGGACTCGCTCATTTACGGAGCATGCGT GGTCTTCACCCTTGGCATGTTCTCCGCCGGCCTCTCGGACCTCAGGCACATGCGAATGACCCGGAGT GTGGACAACGTCCAGTTCCTGCCCTTTCTCACCACGGAAGTCAACAACCTGGGCTGGCTGAGTTATG GGGCTTTGAAGGGAGACGGGATCCTCATCGTCGTCAACACAGTGGGTGTTGTGCTCCTACAGACTGC AACCCTGCTAGGGGTCCTTCTCCTGGGTTATGGCTACTTTTGGCTCCTGGTACCCAACCCTGAGGCC CGGCTTCAGCAGTTGGGCCTCTTCTGCAGTGTCTTCACCATCAGCATGTACCTCTCACCACTGGCTG ACTTGGCTAAGGTGATTCAAACTAAATCAACCCAATGTCTCTCCTACCCACTCACCATTGCTACCCT TCTCACCTCTGCCTCCTGGTGCCTCTATGGGTTTCGACTCAGAGATCCCTATATCATGGTGTCCAAC TTTCCAGGAATCGTCACCAGCTTTATCCGCTTCTGGCTTTTCTGGAAGTACCCCCAGGAGCAAGACA GGAACTACTGGCTCCTGCAAACCTGAGGCTGCTCATCTGACCACTGGGCACCTTAGTGCCAACCTGA ACCAAAGAGACCTCCTTGTTTTATGCTGGG		
	ORF Start: ATG at 21		ORF Stop: TGA at 627

	SEQ ID NO: 60	202 aa	MW at 22754.5kD
NOV15a, CG145122-01 Protein Sequence	MEAGGFLDSLIYGACVVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLTTEVNNLGWLSYGALKGDG ILIVVNTVGVVLLQTATLLGVLLLGYGYFWLLVPNPEARLQQLGLFCSVFTISMYLSPLADLAKVIQ TKSTQCLSYPLTIATLLTSASWCLYGFRLRDPYIMVSNFPGIVTSFIRFWLFWKYPQEQDRNYWLLQ T		

Further analysis of the NOV15a protein yielded the following properties shown in Table 15B.

Table 15B. Protein Sequence Properties NOV15a	
PSort analysis:	0.7300 probability located in plasma membrane; 0.6400 probability located in endoplasmic reticulum (membrane); 0.3880 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 22 and 23

A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15C.

Table 15C. Geneseq Results for NOV15a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV15a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABB90191	Human polypeptide SEQ ID NO 2567 - Homo sapiens,	1..202 66..286	202/221 (91%) 202/221 (91%)	e-112

	286 aa. [WO200190304-A2, 29-NOV-2001]			
AAB75379	Human secreted protein #38 - Homo sapiens, 221 aa. [WO200100806-A2, 04-JAN-2001]	1..202 1..221	202/221 (91%) 202/221 (91%)	e-112
AAE03982	Human gene 43 encoded secreted protein fragment, SEQ ID NO:180 - Homo sapiens, 221 aa. [WO200077022-A1, 21-DEC-2000]	1..202 1..221	202/221 (91%) 202/221 (91%)	e-112
AAB25793	Human secreted protein SEQ ID #105 - Homo sapiens, 221 aa. [WO200037491-A2, 29-JUN-2000]	1..202 1..221	202/221 (91%) 202/221 (91%)	e-112
AAB53433	Human colon cancer antigen protein sequence SEQ ID NO:973 - Homo sapiens, 248 aa. [WO200055351-A1, 21-SEP-2000]	1..202 28..248	202/221 (91%) 202/221 (91%)	e-112

In a BLAST search of public sequence datbases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15D.

Table 15D. Public BLASTP Results for NOV15a

Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BRV3	Stromal cell protein - Homo sapiens (Human), 221 aa.	1..202 1..221	202/221 (91%) 202/221 (91%)	e-112
Q9UHQ3	Stromal cell protein - Homo sapiens (Human), 221 aa.	1..202 1..221	201/221 (90%) 202/221 (90%)	e-112
Q95KW8	Uterine stromal cell protein - Papio anubis (Olive baboon), 221 aa.	1..202 1..221	197/221 (89%) 198/221 (89%)	e-108
Q9UHQ2	Stromal cell protein isoform - Homo sapiens (Human), 179 aa.	1..202 1..179	171/202 (84%) 175/202 (85%)	1e-90
Q9CXK4	Recombination activating gene 1 gene activation - Mus musculus (Mouse), 221 aa.	1..202 1..221	161/221 (72%) 174/221 (77%)	4e-85

PFam analysis predicts that the NOV15a protein contains the domains shown in the Table 15E.

Table 15E. Domain Analysis of NOV15a			
Pfam Domain	**NOV15a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
MtN3_slv	9..79	27/73 (37%) 61/73 (84%)	5.6e-25
MtN3_slv	108..194	35/89 (39%) 77/89 (87%)	1.9e-35

Example 16.

The NOV16 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 16A.

Table 16A. NOV16 Sequence Analysis			
	SEQ ID NO: 61	568 bp	
NOV16a, CG145198-01 DNA Sequence	CCGGCCGGGCC**ATG**GATTCAATGCCTGAGCCCGCGTCCCGCTGTCTTCTGCTTCTTCCCTTGCTGCT GCTGCTGCTGCTGCTGCTGCCGGCCCCGGAGCTGGGCCCGAGCCAGGCCGGAGCTGAGGAGAACGAC TGGGTTCGCCTGCCCAGCAAATGCGAAGTGTGTAAATATGTTGCTGTGGAGCTGAAGTCAGCCTTTG AGGAAACCGGCAAGACCAAGGAGGTGATTGGCACGGGCTATGGCATCCTGGACCAGAAGGCCTCTGG AGTCAAATACACCAAGTCCATTTCAGATCCCCCAGACCAGATGACCTATCTTCCTTCCAGCTCTGAG TCACTTCCCATTGGGACTTGCGGT**TAA**TCGAAGTCACTGAGACCATTTGCAAGAGGCTCCTGGATTA TAGCCTGCACAAGGAGAGGACCGGCAGCAATCGATTTGCCAAGGTTGGATTCGGGATTGTCCTTCAT CCGCTCTGGGGTCAGGCCTGCATGTATCTTAGTGTGTCTGCTGGTGTGAGTGTGATTTGAAGATGAC CACCTGGGATCTTCCCTCATTGCCTCTTCCCT		
	ORF Start: ATG at 12		ORF Stop: TAA at 360

	SEQ ID NO: 62	116 aa	MW at 12441.2kD
NOV16a, CG145198-01 Protein Sequence	MDSMPEPASRCLLLLPLLLLLLLLPAPELGPSQAGAEENDWVRLPSKCEVCKYVAVELKSAFEETG KTKEVIGTGYGILDQKASGVKYTKSISDPPDQMTYLPSSSESLPIGTCG		

	SEQ ID NO: 63	370 bp
NOV16b, 278498076 DNA Sequence	**CACC**GGATCCACCATGGATTCAATGCCTGAGCCCGCGTCCCGCTGTCTTCTGCTTCTTCCCTTGCTG CTGCTGCTGCTGCTGCTGCTGCCGGCCCCGGAGCTGGGCCCGAGCCAGGCCGGAGCTGAGGAGAACG ACTGGGTTCGCCTGCCCAGCAAATGCGAAGTGTGTAAATATGTTGCTGTGGAGCTGAAGTCAGCCTT TGAGGAAACCGGCAAGACCAAGGAGGTGATTGGCACGGGCTATGGCATCCTGGACCAGAAGGCCTCT GGAGTCAAATACACCAAGTCCATTTCAGATCCCCCAGACCAGATGACCTATCTTCCTTCCAGCTCTG AGTCACTTCCCATTGGGACTTGCGGTCTCGAGGGC	
	ORF Start: at 2	ORF Stop: end of sequence

	SEQ ID NO: 64	123 aa	MW at 13086.9kD

NOV16b, 278498076 Protein Sequence	TGSTMDSMPEPASRCLLLLPLLLLLLLLLPAPELGPSQAGAEENDWVRLPSKCEVCKYVAVELKSAF EETGKTKEVIGTGYGILDQKASGVKYTKSISDPPDQMTYLPSSSESLPIGTCGLEG

	SEQ ID NO: 65	274 bp
NOV16c, 278498091 DNA Sequence	CACCGGATCCCCGAGCCAGGCCGGAGCTGAGGAGAACGACTGGGTTCGCCTGCCCAGCAAATGCGAA GTGTGTAAATATGTTGCTGTGGAGCTGAAGTCAGCCTTTGAGGAAACCGGCAAGACCAAGGAGGTGA TTGGCACGGGCTATGGCATCCTGGACCAGAAGGCCTCTGGAGTCAAATACACCAAGTCCATTTCAGA TCCCCCAGACCAGATGACCTATCTTCCTTCCAGCTCTGAGTCACTTCCCATTGGGACTTGCGGTCTC GAGGGC	
	ORF Start: at 2	ORF Stop: end of sequence

	SEQ ID NO: 66	91 aa	MW at 9647.7kD
NOV16c, 278498091 Protein Sequence	TGSPSQAGAEENDWVRLPSKCEVCKYVAVELKSAFEETGKTKEVIGTGYGILDQKASGVKYTKSISD PPDQMTYLPSSSESLPIGTCGLEG		

	SEQ ID NO: 67	1596 bp	
NOV16d, CG145198-02 DNA Sequence	CCCAGGCCCAGACGCAGGCTTCTTCTCCTCGGGTCTTGGTCCTGCATCCTCTCTCTCCCAGAGCCTC CGTTAGGGGGTGGAAAGGACTTTGCCATAGGTCGCTGAGGCCACCATCTGCTCTCTTACTGGCCAA GGGCGTAAAAAGATAGTCCTCCCATTAGCTAGAGAGCAAACCCCAGAAAGCCTATTGGCTGCGCCGT CCGCGGGCCTTGGTCCGCTTTGAAGGCGGGCTGCGGCTGCGAGAGGAGGGCGGGCGGGAGGCTAGCT GTTGTCGTGGTTGCTCGGAGGCACGTGTGCAGTCCCGGAAGCGGCGAGGGGAAACTGCTCCGCGCGC GCCGCGGGAGGAGGAACCGCCCGGTCCTTTAGGGTCCGGGCCCGGCCGGGCC**ATG**GATTCAATGCCT GAGCCCGCGTCCCGCTGTCTTCTGCTTCTTCCCTTGCTGCTGCTGCTGCTGCTGCTGCTGCCGGCCC CGGAGCTGGGCCCGAGCCAGGCCGGAGCTGAGGAGAACGACTGGGTTCGCCTGCCCAGCAAATGCGA AGGGACTTGCGGT**TAA**TCGAAGTCACTGAGACCATTTGCAAGAGGCTCCTGGATTATAGCCTGCACA AGGAGAGGACCGGCAGCAATCGATTTGCCAAGGGCATGTCAGAGACCTTTGAGACATTACACAACCT GGTACACAAAGGGGTCAAGGTGGTGATGGACATCCCCTATGAGCTGTGGAACGAGACTTCTGCAGAG GTGGCTGACCTCAAGAAGCAGTGTGATGTGCTGGTGGAAGAGTTTGAGGAGGTGATCGAGGACTGGT ACAGGAACCACCAGGAGGAAGACCTGACTGAATTCCTCTGCGCCAACCACGTGCTGAAGGGAAAAGA CACCAGTTGCCTGGCAGAGCAGTGGTCCGGCAAGAAGGGAGACACAGCTGCCCTGGAGGGAAGAAG TCCAAGAAGAAGAGCAGCAGGGCCAAGGCAGCAGGCGGCAGGAGTAGCAGCAGCAAACAAAGGAAGG AGCTGGGTGGCCTTGAGGGAGACCCCAGCCCCGAGGAGGATGAGGGCATCCAGAAGGCATCCCCTCT CACACACAGCCCCCCTGATGAGCTCTGAGCCCACCCAGCATCCTCTGTCCTGAGACCCCTGATTTTG AAGCTGAGGAGTCAGGGGCATGGCTCTGGCAGGCCGGGATGGCCCCGCAGCCTTCAGCCCCTCCTTG CCTTGGCTGTGCCCTCTTCTGCCAAGGAAAGACACAAGCCCCAGGAAGAACTCAGAGCCGTCATGGG TAGCCCACGCCGTCCTTTCCCCTCCCCAAGTGTTTCTCTCCTGACCCAGGGTTCAGGCAGGCCTTGT GGTTTCAGGACTGCAAGGACTCCAGTGTGAACTCAGGAGGGGCAGGTGTCAGAACTGGGCACCAGGA CTGGAGCCCCCTCCGGAGACCAAACTCACCATCCCTCAGTCCTCCCCAACAGGGTACTAGGACTGCA GCCCCCTGTAGCTCCTCTCTGCTTACCCCTCCTGTGGACACCTTGCACTCTGCCTGGCCCTTCCCAG AGCCCAAAGAGTAAAAATGTTCTGGTTCTGAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 388		ORF Stop: TAA at 550

	SEQ ID NO: 68	54 aa	MW at 5772.7kD
NOV16d, CG145198-02 Protein Sequence	MDSMPEPASRCLLLLPLLLLLLLLLPAPELGPSQAGAEENDWVRLPSKCEGTCG		

	SEQ ID NO: 69	901 bp	
NOV16e, CG145198-03 DNA Sequence	GGAGGAGGAACCGCCCGGTCCTTTAGGGTCCGGGCCCGGCCGGGCC**ATG**GATTCAATGCCTGAGCCC GCGTCCCGCTGTCTTCTGCTTCTTCCCTTGCTGCTGCTGCTGCTGCTGCTGCTGCCGGCCCCGGAGC TGGGCCCGAGCCAGGCCGGAGCTGAGGAGAACGACTGGGTTCGCCTGCCCAGCAAATGCGAAGTGTG TAAATATGTTGCTGTGGAGCTGAAGTCAGCCTTTGAGGAAACCGGCAAGACCAAGGAGGTGATTGGC		

	ACGGGCTATGGCATCCTGGACCAGAAGGCCTCTGGAGTCAAATACACCAAGTCGGACTTGCGGTTAA TCGAAGTCACTGAGACCATTTGCAAGAGGCTCCTGGATTATAGCCTGCACAAGGAGAGGACCGGCAG CAATCGATTTGCCAAGGGCATGTCAGAGACCTTTGAGACATTACACAACCTGGTACACAAAGGGGTC AAGGTGGTGATGGACATCCCCTATGAGCTGTGGAACGAGACTTCTGCAGAGGTGGCTGACCTCAAGA AGCAGTGTGATGTGCTGGTGGAAGAGTTTGAGGAGGTGATCGAGGACTGGTACAGGAACCACCAGGA GGAAGACCTGACTGAATTCCTCTGCGCCAACCACGTGCTGAAGGGAAAAGACACCAGTTGCCTGGCA GAGCAGTGGTCCGGCAAGAAGGGAGACACAGCTGCCCTGGGAGGGAAGAAGCCCAAGAAGAAGAGCA GCAGGGCCAAGGCAGCAGGCGGCAGGAGTAGCAGCAGCAAACAAAGGAAGGAGCTGGGTGGCCTTGA GGGAGACCCCAGCCCCGAGGAGGATGAGGGCATCCAGAAGGCATCCCCTCTCACACACAGCCCCCCT GATGAGCTCTGAGCCCACCCAGCATCCTCT		
	ORF Start: ATG at 47		ORF Stop: TGA at 881

	SEQ ID NO: 70	278 aa	MW at 30757.7kD
NOV16e, CG145198-03 Protein Sequence	MDSMPEPASRCLLLLPLLLLLLLLPAPELGPSQAGAEENDWVRLPSKCEVCKYVAVELKSAFEETG KTKEVIGTGYGILDQKASGVKYTKSDLRLIEVTETICKRLLDYSLHKERTGSNRFAKGMSETFETLH NLVHKGVKVVMDIPYELWNETSAEVADLKKQCDVLVEEFEEVIEDWYRNHQEEDLTEFLCANHVLKG KDTSCLAEQWSGKKGDTAALGGKKPKKKSSRAKAAGGRSSSSKQRKELGGLEGDPSPEEDEGIQKAS PLTHSPPDEL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 16B.

Table 16B. Comparison of NOV16a against NOV16b through NOV16e.

Protein Sequence	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV16b	1..116 5..120	101/116 (87%) 101/116 (87%)
NOV16c	32..116 4..88	85/85 (100%) 85/85 (100%)
NOV16d	1..50 1..50	35/50 (70%) 35/50 (70%)
NOV16e	1..92 1..92	77/92 (83%) 77/92 (83%)

Further analysis of the NOV16a protein yielded the following properties shown in Table 16C.

Table 16C. Protein Sequence Properties NOV16a

PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 32 and 33

A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 16D.

Table 16D. Geneseq Results for NOV16a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV16a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABP41913	Human ovarian antigen HVVBT41, SEQ ID NO:3045 - Homo sapiens, 353 aa. [WO200200677-A1, 03-JAN-2002]	1..92 76..167	92/92 (100%) 92/92 (100%)	2e-47
AAU02499	Human trinucleotide repeat protein (TRP) - Homo sapiens, 278 aa. [WO200130798-A1, 03-MAY-2001]	1..92 1..92	92/92 (100%) 92/92 (100%)	2e-47
AAU12239	Human PRO4409 polypeptide sequence - Homo sapiens, 278 aa. [WO200140466-A2, 07-JUN-2001]	1..92 1..92	92/92 (100%) 92/92 (100%)	2e-47
AAW78312	Fragment of human secreted protein encoded by gene 67 - Homo sapiens, 277 aa. [WO9856804-A1, 17-DEC-1998]	1..92 1..91	82/92 (89%) 83/92 (90%)	3e-39
AAU02498	Murine trinucleotide repeat protein (TRP) - Mus sp, 276 aa. [WO200130798-A1, 03-MAY-2001]	1..92 1..92	78/92 (84%) 80/92 (86%)	4e-37

In a BLAST search of public sequence datbases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16E.

Table 16E. Public BLASTP Results for NOV16a				
Protein Accession Number	**Protein/Organism/Length**	**NOV16a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**

Q9BT09	Hypothetical 30.7 kDa protein (Unknown) (Protein for MGC:4122) (Protein for MGC:1220) (DJ475N16.1) (CTG4A) - Homo sapiens (Human), 278 aa.	1..92 1..92	92/92 (100%) 92/92 (100%)	4e-47
O15412	CTG4a - Homo sapiens (Human), 143 aa.	1..92 1..92	92/92 (100%) 92/92 (100%)	4e-47
Q9DAU1	1600025D17Rik protein (Putative retinoic acid-regulated protein) (RIKEN cDNA 1600025D17 gene) - Mus musculus (Mouse), 276 aa.	1..92 1..92	78/92 (84%) 80/92 (86%)	1e-36
CAC39850	Sequence 345 from Patent EP1067182 - Homo sapiens (Human), 248 aa.	19..76 8..65	24/58 (41%) 35/58 (59%)	6e-06
Q8WUN9	Hypothetical 29.4 kDa protein - Homo sapiens (Human), 257 aa (fragment).	19..76 19..76	24/58 (41%) 35/58 (59%)	6e-06

PFam analysis predicts that the NOV16a protein contains the domains shown in the Table 16F.

Table 16F. Domain Analysis of NOV16a			
Pfam Domain	**NOV16a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**

Example 17.

The NOV17 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis			
	SEQ ID NO: 71	862 bp	
NOV17a, CG145286-01 DNA Sequence	CCCCTCCCATTTGCCTGTCCTGGTCAGGCCCCCACCCCCCTTCCCACCTGACCAGCCATGGGGGCTG CGGTGTTTTTCGGCTGCACTTTCGTCGCGTTCGGCCCGGCCTTCGCGCTTTTCTTGATCACTGTGGC TGGGGACCCGCTTCGCGTTATCATCCTGGTCGCAGGGGCATTTTTCTGGCTGGTCTCCCTGCTCCTG GCCTCTGTGGTCTGGTTCATCTTGGTCCATGTGACCGACCGGTCAGATGCCCGGCTCCAGTACGGCC TCCTGATTTTTGGTGCTGCTGTCTCTGTCCTTCTACAGGAGGTGTTCCGCTTTGCCTACTACAAGCT GCTTAAGAAGGCAGATGAGGGGTTAGCATCGCTGAGTGAGGACGGAAGATCACCCATCTCCATCCGC CAGATGGCCTATGGTGTGGTTGGGATCCATGGAGACTCACCCTATTACTTCCTGACTTCAGCCTTTC TGACAGCAGCCATTATCCTGCTCCATACCTTTTGGGGAGTTGTGTTCTTTGATGCCTGTGAGAGGAG ACGGTACTGGGCTTTGGGCCTGGTGGTTGGGAGTCACCTACTGACATCGGGACTGACATTCCTGAAC		

	CCCTGGTATGAGGCCAGCCTGCTGCCCATCTATGCAGTCACTGTTTCCATGGGGCTCTGGGCCTTCA TCACAGCTGGAGGGTCCCTCCGAAGTATTCAGCGCAGCCTCTTGTGCCGACGGCAGGAGGACAGTCG GGTGATGGTGTATTCTGCCCTGCGCATCCCACCCGAGGACTGAGGGAACCTAGGGGGGACCCCTGGG CCTGGGGTGCCCTCCTGATGTCCTCGCCCTGTATTTCTCCATCTCCAGTTCTGGACAG		
	ORF Start: ATG at 58		ORF Stop: TGA at 778

	SEQ ID NO: 72	240 aa	MW at 26566.8kD
NOV17a, CG145286-01 Protein Sequence	MGAAVFFGCTFVAFGPAFALFLITVAGDPLRVIILVAGAFFWLVSLLLASVVWFILVHVTDRSDARL QYGLLIFGAAVSVLLQEVFRFAYYKLLKKADEGLASLSEDGRSPISIRQMAYGVVGIHGDSPYYFLT SAFLTAAIILLHTFWGVVFFDACERRRYWALGLVVGSHLLTSGLTFLNPWYEASLLPIYAVTVSMGL WAFITAGGSLRSIQRSLLCRRQEDSRVMVYSALRIPPED		

	SEQ ID NO: 73	942 bp	
NOV17b, CG145286-02 DNA Sequence	CCTTCCCCTCCCATTTGCCTGTCCTGGTCAGGCCCCCCACCCCCCTTCCCACCTGACCAGCCATGGG GGCTGCGGTGTTTTTCGGCTGCACTTTCGTCGCGTTCGGCCCGGCCTTCGCGCTTTTCTTGATCACT GTGGCTGGGGACCCGCTTCGCGTTATCATCCTGGTCGCAGGGGCATTTTCCTGGCTGGTCTCCCTGC TCCTGGCCTCTGTGGTCTGGTTCATCTTGGTCCATGTGACCGACCGGTCAGATGCCCGGCTCCAGTA CGGCCTCCTGATTTTTGGTGCTGCTGTCTCTGTCCTTCTACAGGAGGTGTTCCGCTTTGCCTACTAC AAGCTGCTTAAGAAGGCAGATGAGGGGTTAGCATCGCTGAGTGAGGACGGAAGATCACCCATCTCCA TCCGCCAGATGGCCTATGTTTCTGGTCTCTCCTTCGGTATCATCAGTGGTGTCTTCTCTGTTATCAA TATTTTGGCTGATGCACTTGGGCCAGGTGTGGTTGGGATCCATGGAGACTCACCCTATTACTTCCTG ACTTCAGCCTTTCTGACAGCAGCCATTATCCTGCTCCATACCTTTTGGGGAGTTGTGTTCTTTGATG CCTGTGAGAGGAGACGGTACTGGGCTTTGGGCCTGGTGGTTGGGAGTCACCTACTGACATCGGGACT GACATTCCTGAACCCCTGGTATGAGGCCAGCCTGCTGCCCATCTATGCAGTCACTGTTTCCATGGGG CTCTGGGCCTTCATCACAGCTGGAGGGTCCCTCCGAAGTATTCAGCGCAGCCTCTTGTGCCGACGGC AGGAGGACAGTCGGGTGATGGTGTATTCTGCCCTGCGCATCCCACCCGAGGACTGAGGGAACCTAGG GGGGACCCCTGGGCCTGGGGTGCCCTCCTGATGTCCTCGTCCTGTATTTCTCCATCTCCAGTTCTGG ACAG		
	ORF Start: ATG at 63		ORF Stop: TGA at 858

	SEQ ID NO: 74	265 aa	MW at 28935.5kD
NOV17b, CG145286-02 Protein Sequence	MGAAVFFGCTFVAFGPAFALFLITVAGDPLRVIILVAGAFSWLVSLLLASVVWFILVHVTDRSDARL QYGLLIFGAAVSVLLQEVFRFAYYKLLKKADEGLASLSEDGRSPISIRQMAYVSGLSFGIISGVFSV INILADALGPGVVGIHGDSPYYFLTSAFLTAAIILLHTFWGVVFFDACERRRYWALGLVVGSHLLTS GLTFLNPWYEASLLPIYAVTVSMGLWAFITAGGSLRSIQRSLLCRRQEDSRVMVYSALRIPPED		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 17B.

Table 17B. Comparison of NOV17a against NOV17b.		
Protein Sequence	**NOV17a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV17b	1..240 1..265	224/265 (84%) 224/265 (84%)

Further analysis of the NOV17a protein yielded the following properties shown in Table 17C.

Table 17C. Protein Sequence Properties NOV17a

PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 63 and 64

A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17D.

Table 17D. Geneseq Results for NOV17a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB65248	Human PRO1141 (UNQ579) protein sequence SEQ ID NO:303 - Homo sapiens, 247 aa. [WO200073454-A1, 07-DEC-2000]	1..221 1..246	220/246 (89%) 221/246 (89%)	e-120
AAB94784	Human protein sequence SEQ ID NO:15888 - Homo sapiens, 247 aa. [EP1074617-A2, 07-FEB-2001]	1..221 1..246	220/246 (89%) 221/246 (89%)	e-120
AAM93680	Human polypeptide, SEQ ID NO: 3574 - Homo sapiens, 247 aa. [EP1130094-A2, 05-SEP-2001]	1..221 1..246	220/246 (89%) 221/246 (89%)	e-120
AAU29137	Human PRO polypeptide sequence #114 - Homo sapiens, 247 aa. [WO200168848-A2, 20-SEP-2001]	1..221 1..246	220/246 (89%) 221/246 (89%)	e-120
AAY57881	Human transmembrane protein HTMPN-5 - Homo sapiens, 247 aa. [WO9961471-A2, 02-DEC-1999]	1..221 1..246	220/246 (89%) 221/246 (89%)	e-120

In a BLAST search of public sequence datbases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17E.

Table 17E. Public BLASTP Results for NOV17a

Protein Accession Number	Protein/Organism/Length	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96BI3	Hypothetical 29.0 kDa protein (CGI-78 protein) - Homo sapiens (Human), 265 aa.	1..240 1..265	240/265 (90%) 240/265 (90%)	e-131
Q9BVG0	Similar to CGI-78 protein - Homo sapiens (Human), 265 aa.	1..240 1..265	239/265 (90%) 240/265 (90%)	e-131
Q8R1T3	CGI-78 protein - Mus musculus (Mouse), 265 aa.	1..240 1..265	238/265 (89%) 239/265 (89%)	e-130
Q969R6	CGI-78 protein - Homo sapiens (Human), 247 aa.	1..221 1..246	220/246 (89%) 221/246 (89%)	e-119
CAC39761	Sequence 159 from Patent EP1067182 - Homo sapiens (Human), 247 aa.	1..221 1..246	219/246 (89%) 220/246 (89%)	e-118

PFam analysis predicts that the NOV17a protein contains the domains shown in the Table 17F.

Table 17F. Domain Analysis of NOV17a			
Pfam Domain	**NOV17a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**

Example 18.

The NOV18 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis		
	SEQ ID NO: 75	644 bp
NOV18a, CG145650-01 DNA Sequence	GTAATTTACCACCATCTTTGGTTCCTGTTTATAAGATGTTTTAAGAAAGATTTGAAACAGATTTTCT GAAGAAAGCAGAAGCTCTCTTCCCATTATGACTTCGGAAATCACTTATGCTGAAGTGAGGTTCAAAA ATGAATTCAAGTCCTCAGGCATCAACACAGCCTCTTCTGCAGAGACAGCCTGGAGCTGTTGCCCAAA GAATTGGAAGTCATTTAGTTCCAACTGCTACTTTATTTCTACTGAATCAGCATCTTGGCAAGACAGT GAGAAGGACTGTGCTAGAATGGAGGCTCACCTGCTGGTGATAAACACTCAAGAAGAGCAGGATTTCA TCTTCCAGAATCTGCAAGAAGAATCTGCTTATTTTTTGGGGCTCTCAGATCCAGAAGGTCAGCGACA TTGGCAATGGGTTGATCAGACGCCATACAATGAAAGTTCCACATTCTGGCATCCACGTGAGCCCAGT GATCCCAATGAGCGCTGCGTTGTGCTAAATTTTCGTAAATCACCCAAAAGATGGGGCTGGAATGATG TTAATTGTCTTGGTCCTCAAAGGTCAGTTTGTGAGATGATGAAGATCCACTTATGAACTGAACATTC TCCATGAACAGGTGGTTGGATTGGTATCTGTCATTGTAGGG

	ORF Start: ATG at 95		ORF Stop: TGA at 590

	SEQ ID NO: 76	165 aa	MW at 19294.2kD
NOV18a, CG145650-01 Protein Sequence	MTSEITYAEVRFKNEFKSSGINTASSAETAWSCCPKNWKSFSSNCYFISTESASWQDSEKDCARMEA HLLVINTQEEQDFIFQNLQEESAYFLGLSDPEGQRHWQWVDQTPYNESSTFWHPREPSDPNERCVVL NFRKSPKRWGWNDVNCLGPQRSVCEMMKIHL		

	SEQ ID NO: 77	763 bp	
NOV18b, CG145650-02 DNA Sequence	GTAATTTACCACCATGTTTGGTTCCTGTTTATAAGATGTTTTAAGAAAGATTTGAAACAGATTTTCT GAAGAAAGCAGAAGCTCTCTTCCCATTATGACTTCGGAAATCACTTATGCTGAAGTGAGGTTCAAAA ATGAATTCAAGTCCTCAGGCATCAACACAGCCTCTTCTGCAGCTTCCAAGGAGAGGACTGCCCCTCT CAAAAGTAATACCGGATTCCCCAAGCTGCTTTGTGCCTCACTGTTGATATTTTTCCTGCTATTGGCA ATCTCATTCTTTATTGCTTTTGTCATTTTCTTTCAAAAATATTCTCAGCTTCTTGAAAAAAAGACTA CAAAAGAGCTGGTTCATACAACATTGGAGTGTGTGAAAAAAAATATGCCCGTGGAAGAGACAGCCTG GAGCTGTTGCCCAAAGAATTGGAAGTCATTTAGTTCCAACTGCTACTTTATTTCTACTGAATCAGCA TCTTGGCAAGACAGTGAGAAGGACTGTGCTAGAATGGAGGCTCACCTGCTGGTGATAAACACTCAAG AAGAGCAGGATTTCATCTTCCAGAATCTGCAAGAAGAATCTGCTTATTTTGTGGGGCTCTCAGATCC AGAAGGTCAGCGACATTGGCAATGGGTTGATCAGACACCATACAATGATGTTAATTGTCTTGGTCCT CAAAGGTCAGTTTGTGAGATGATGAAGATCCACTTATGAACTGAACATTCTCCCATGAAACAGGTGG TTGGATTGGTATCTGTCATTGTAGGG		
	ORF Start: ATG at 95		ORF Stop: TGA at 707

	SEQ ID NO: 78	204 aa	MW at 23462.5kD
NOV18b, CG145650-02 Protein Sequence	MTSEITYAEVRFKNEFKSSGINTASSAASKERTAPLKSNTGFPKLLCASLLIFFLLLAISFFIAFVI FFQKYSQLLEKKTTKELVHTTLECVKKNMPVEETAWSCCPKNWKSFSSNCYFISTESASWQDSEKDC ARMEAHLLVINTQEEQDFIFQNLQEESAYFVGLSDPEGQRHWQWVDQTPYNDVNCLGPQRSVCEMMK IHL		

	SEQ ID NO: 79	1308 bp	
NOV18c, CG145650-03 DNA Sequence	CTCACTATACTGGTCCTGAGGAAAGGGCTTCTGTGAACTGCGGTTTTTAGTTTTTATTGTGGTTCTT AGTTCTCATGAGACCCCTCTTGAGGATATGTGCCTATCTGGTGCCTCTGCTCTCCACTAGTTGAGTG AAAGGAAGGAGGTAATTTACCACCATGTTTGGTTCCTGTTTATAAGATGTTTTAAGAAAGATTTGAA ACAGATTTTCTGAAGAAAGCAGAAGCTCTCTTCCCATTATGACTTCGGAAATCACTTATGCTGAAGT GAGGTTCAAAAATGAATTCAAGTCCTCAGGCATCAACACAGCCTCTTCTGCAGCTTCCAAGGAGAGG ACTGCCCCTCTCAAAAGTAATACCGGATTCCCCAAGCTGCTTTGTGCCTCACTGTTGATATTTTTCC TGCTATTGGCAATCTCATTCTTTATTGCTTTTGTCATTTTCTTTCAAAAATATTCTCAGCTTCTTGA AAAAAAGACTACAAAAGAGCTGGTTCATACAACATTGGAGTGTGTGAAAAAAAATATGCCCGTGGAA GAGACAGCCTGGAGCTGTTGCCCAAAGAATTGGAAGTCATTTAGTTCCAACTGCTACTTTATTTCTA CTGAATCAGCATCTTGGCAAGACAGTGAGAAGGACTGTGCTAGAATGGAGGCTCACCTGCTGGTGAT AAACACTCAAGAAGAGCAGGATTTCATCTTCCAGAATCTGCAAGAAGAATCTGCTTATTTTGTGGGG CTCTCAGATCCAGAAGGTCAGCGACATTGGCAATGGGTTGATCAGACACCATACAATGAAAGTTCCA CATTCTGGCATCCACGTGAGCCCAGTGATCCCAATGAGCGCTGCGTTGTGCTAAATTTTCGTAAATC ACCCAAAAGATGGGGCTGGAATGATGTTAATTGTCTTGGTCCTCAAAGGTCCAGTTTGTGAGATGAT GAAGATCCACTTATGAACTGAACATTCTCCATGAACAGGTGGTTGGATTGGTATCTGTCATTGTAGG GATAGATAATAAGCTCTTCTTATTCATGTGTAAGGGAGGTCCATAGAATTTAGGTGGTCTGTCAACT ATTCTACTTATGAGAGAATTGGTCTGTACATTGACTGATTCACTTTTTCATAAAGTGAGCATTTATT GAGCATTTTTTCATGTGCCAGAGCCTGTACTGGAGGCCCCCATTGTGCACACATGGAGAGAACATGA GTCTCTCTTAATTTTTATCTGGTTGCTAAAGAATTATTTACCAATAAAATTATATGATGTGGTGAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 240		ORF Stop: TGA at 930

	SEQ ID NO: 80	230 aa	MW at 26602.8kD
NOV18c, CG145650-03 Protein Sequence	MTSEITYAEVRFKNEFKSSGINTASSAASKERTAPLKSNTGFPKLLCASLLIFFLLLAISFFIAFVI FFQKYSQLLEKKTTKELVHTTLECVKKNMPVEETAWSCCPKNWKSFSSNCYFISTESASWQDSEKDC ARMEAHLLVINTQEEQDFIFQNLQEESAYFVGLSDPEGQRHWQWVDQTPYNESSTFWHPREPSDPNE RCVVLNFRKSPKRWGWNDVNCLGPQRSSL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 18B.

Table 18B. Comparison of NOV18a against NOV18b and NOV18c.		
Protein Sequence	**NOV18a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV18b	28..165 100..204	104/138 (75%) 105/138 (75%)
NOV18c	28..156 100..228	128/129 (99%) 129/129 (99%)

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

Table 18C. Protein Sequence Properties NOV18a	
PSort analysis:	0.6868 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18D.

Table 18D. Geneseq Results for NOV18a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV18a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABP48034	Human polypeptide SEQ ID NO 464 - Homo sapiens, 243 aa. [US2002042386-A1, 11-APR-2002]	28..165 106..243	137/138 (99%) 138/138 (99%)	3e-84
ABP47873	Human polypeptide SEQ ID NO 303 - Homo sapiens, 246 aa. [US2002042386-A1, 11-APR-2002]	28..165 109..246	137/138 (99%) 138/138 (99%)	3e-84

AAU98014	Human dendritic cell immunoreceptor AJ133532 - Homo sapiens, 237 aa. [WO200232958-A2, 25-APR-2002]	28..165 100..237	137/138 (99%) 138/138 (99%)	3e-84
ABB90277	Human polypeptide SEQ ID NO 2653 - Homo sapiens, 237 aa. [WO200190304-A2, 29-NOV-2001]	28..165 100..237	137/138 (99%) 138/138 (99%)	3e-84
AAU19814	Human novel extracellular matrix protein, Seq ID No 464 - Homo sapiens, 243 aa. [WO200155368-A1, 02-AUG-2001]	28..165 106..243	137/138 (99%) 138/138 (99%)	3e-84

In a BLAST search of public sequence datbases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a

Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H2Z9	C-type lectin DDB27 short form - Homo sapiens (Human), 204 aa.	1..165 1..204	163/204 (79%) 165/204 (79%)	8e-93
Q9UMR7	Dendritic cell immunoreceptor - Homo sapiens (Human), 237 aa.	28..165 100..237	137/138 (99%) 138/138 (99%)	9e-84
Q9UI34	C-type lectin superfamily 6 - Homo sapiens (Human), 237 aa.	28..165 100..237	137/138 (99%) 138/138 (99%)	9e-84
Q9NS33	HDCGC13P - Homo sapiens (Human), 237 aa.	28..165 100..237	136/138 (98%) 137/138 (98%)	3e-83
Q8WXW9	Fc-epsilon receptor III - Homo sapiens (Human), 230 aa.	28..156 100..228	128/129 (99%) 129/129 (99%)	5e-78

PFam analysis predicts that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a

Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lectin_c	51..160	34/127 (27%) 85/127 (67%)	5.8e-28

Example 19.

The NOV19 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis			
	SEQ ID NO: 81	661 bp	
NOV19a, CG145836-01 DNA Sequence	CTCCTGTAACCCTCCTCCAGG**ATG**AACCACCTGCCAGAAGACATGGAGAACGCTCTCACCGGGAGCC AGAGCTCCCATGCTTCTCTGCGCAATATCCATTCCATCAACCCCACACAACTCATGGCCAGGATTGA GTCCTATGAAGGAAGGGAAAAGAAAGGCATATCTGATGTCAGGAGGACTTTCTGTTTGTTTGTCACC TTTGACCTCTTATTCGTAACATTACTGTGGATAATAGAGTTAAATGTGAATGGAGGCATTGAGAACA CATTAGAGAAGGAGGTGATGCAGTATGACTACTATTCTTCATATTTTGATATATTTCTTCTGGCAGT TTTTCGATTTAAAGTGTTAATACTTGCATATGCTGTGTGCAGACTGCCCATCATTTCATTCATCCTT GCCTGGATTGAGACGTGGTTCCTGGATTTCAAAGTGTTACCTCAAGAAGCAGAAGAAGAAAACAGAC TCCTGATAGTTCAGGATGCTTCAGAGAGGGCAGCACTTATACCTGGTGGTCTTTCTGATGGTCAGTT TTATTCCCCTCCTGAATCCGAAGCAGGATCTGAAGAAGCTGAAGAAAAACAGGACAGTGAGAAACCA CTTTTAGAACTA**TGA**GTACTACTTTTGTTAAATGTGAAAAACCCTCACAGAAAGTCAT		
	ORF Start: ATG at 22		ORF Stop: TGA at 616

	SEQ ID NO: 82	198 aa	MW at 22691.5kD
NOV19a, CG145836-01 Protein Sequence	MNHLPEDMENALTGSQSSHASLRNIHSINPTQLMARIESYEGREKKGISDVRRTFCLFVTFDLLFVT LLWIIELNVNGGIENTLEKEVMQYDYYSSYFDIFLLAVFRFKVLILAYAVCRLPIISFILAWIETWF LDFKVLPQEAEEENRLLIVQDASERAALIPGGLSDGQFYSPPESEAGSEEAEEKQDSEKPLLEL		

	SEQ ID NO: 83	769 bp	
NOV19b, CG145836-02 DNA Sequence	CTCCTGTAACCCTCCTCCAGG**ATG**AACCACCTGCCAGAAGACATGGAGAACGCTCTCACCGGGAGCC AGAGCTCCCATGCTTCTCTGCGCAATATCCATTCCATCAACCCCACACAACTCATGGCCAGGATTGA GTCCTATGAAGGAAGGGAAAAGAAAGGCATATCTGATGTCGGGAGGACTTTCTGTTTGTTTGTCACC TTTGACCTCTTATTCGTAACATTACTGTGGATAATAGAGTTAAATGTGAATGGAGGCATTGAGAACA CATTAGAGAAGGAGGTGATGCAGTATGACTACTATTCTTCATATTTTGATATATTTCTTCTGGCAGT TTTTCGATTTAAAGTGTTAATACTTGCATATGCTGTGTGCAGACTGCGCCATTGGTGGGCAATAGCG TTGACAACGGCAGTGACCAGTGCCTTTTTACTAGCAAAAGTGATCCTTTCGAAGCTTTTCTCTCAAG GGGCTTTTGGCTATGTGCTGCCCATCATTTCATTCATCCTTGCCTGGATTGAGACGTGGTTCCTGGA TTTCAAAGTGTTACCTCAAGAAGCAGAAGAAGAAAACAGACTCCTGATAGTTCAGGATGCTTCAGAG AGGGCAGCACTTATACCTGGTGGTCTTTCTGATGGTCAGTTTTATTCCCCTCCTGAATCCGAAGCAG GATCTGAAGAAGCTGAAGAAAAACAGGACAGTGAGAAACCACTTTTAGAACTA**TGA**GTACTACTTTT GTTAAATGTGAAAAACCCTCACAGAAAGTCAT		
	ORF Start: ATG at 22		ORF Stop: TGA at 724

	SEQ ID NO: 84	234 aa	MW at 26555.1kD
NOV19b, CG145836-02 Protein Sequence	MNHLPEDMENALTGSQSSHASLRNIHSINPTQLMARIESYEGREKKGISDVGRTFCLFVTFDLLFVT LLWIIELNVNGGIENTLEKEVMQYDYYSSYFDIFLLAVFRFKVLILAYAVCRLRHWWAIALTTAVTS AFLLAKVILSKLFSQGAFGYVLPIISFILAWIETWFLDFKVLPQEAEEENRLLIVQDASERAALIPG GLSDGQFYSPPESEAGSEEAEEKQDSEKPLLEL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 19B.

Table 19B. Comparison of NOV19a against NOV19b.		
Protein Sequence	**NOV19a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV19b	1..198 1..234	167/234 (71%) 167/234 (71%)

Further analysis of the NOV19a protein yielded the following properties shown in Table 19C.

Table 19C. Protein Sequence Properties NOV19a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 3 and 4

A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 19D.

Table 19D. Geneseq Results for NOV19a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV19a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM39930	Human polypeptide SEQ ID NO 3075 - Homo sapiens, 216 aa. [WO200153312-A1, 26-JUL-2001]	1..198 1..216	198/216 (91%) 198/216 (91%)	e-107
ABB84847	Human PRO1864 protein sequence SEQ ID NO:62 - Homo sapiens, 234 aa. [WO200200690-A2, 03-JAN-2002]	1..198 1..234	198/234 (84%) 198/234 (84%)	e-105
ABB95453	Human angiogenesis related protein PRO1864 SEQ ID NO: 62 - Homo sapiens, 234 aa. [WO200208284-A2, 31-JAN-2002]	1..198 1..234	198/234 (84%) 198/234 (84%)	e-105

AAB87532	Human PRO1864 - Homo sapiens, 234 aa. [WO200116318-A2, 08-MAR-2001]	1..198 1..234	198/234 (84%) 198/234 (84%)	e-105
AAM41716	Human polypeptide SEQ ID NO 6647 - Homo sapiens, 238 aa. [WO200153312-A1, 26-JUL-2001]	1..198 5..238	198/234 (84%) 198/234 (84%)	e-105

In a BLAST search of public sequence datbases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19E.

Table 19E. Public BLASTP Results for NOV19a

Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O95772	H_NH1021A08.1 protein (Unknown) (Protein for MGC:14607) (Similar to steroidogenic acute regulatory protein related) - Homo sapiens (Human), 234 aa.	1..198 1..234	198/234 (84%) 198/234 (84%)	e-105
Q99J63	Similar to RIKEN cDNA 0610035N01 gene - Mus musculus (Mouse), 235 aa.	1..198 1..235	186/235 (79%) 191/235 (81%)	1e-96
Q9DCI3	0610035N01Rik protein - Mus musculus (Mouse), 235 aa.	1..198 1..235	185/235 (78%) 190/235 (80%)	3e-96
Q9D356	6530409L22Rik protein - Mus musculus (Mouse), 272 aa.	30..193 39..238	145/200 (72%) 151/200 (75%)	2e-73
Q61542	MLN 64 protein (ES 64 protein) (StarD3) - Mus musculus (Mouse), 446 aa.	7..193 11..229	105/224 (46%) 133/224 (58%)	1e-45

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19F.

Table 19F. Domain Analysis of NOV19a

Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 20.

The NOV20 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 20A.

Table 20A. NOV20 Sequence Analysis

	SEQ ID NO: 85	3157 bp	
NOV20a, CG145978-01 DNA Sequence	GCGGCAGTAGCAGCCATGCTGCCCTTTCTGCTGGCCACACTGGGCACCACAGCCCTCAACAACAGCA ACCCCAAGGACTACTGCTACAGCGCCCGCATCCGCAGCACTGTCCTGCAGGGCCTGCCCTTTGGGGG CGTCCCCACCGTGCTGGCTCTCGACTTCATGTGCTTCCTTTTCCCTCAGGCACTGCTGTTCTTATTC TCTATCCTCCGGAAGGTGGCCTGGGACTATGGGCGGCTGGCCTTGGTGACAGATGCAGACAGCCATG ACCGGTATGAGCGTCTCACCTCTGTCTCCAGCTCCGTTGACTTTGACCAAAGGGACAATGTGGGTTT CTGTTCCTGGCTGACAGCCATCTTCAGGATAGATGATGAGATCCGGGACAAATGTGGGGGCGATGCC GTGCACTACCTGTCCTTTCAGCGGCACATCATCGGGCTGCTGGTGGTTGTGGGCGTCCTCTCCGTAG GCATCGTGCTGCCTGTCAACTTCTCAGGGGACCTGCTGGAGAACAATGCCTACAGCTTTGGGAGAAC CACCATTGCCAACTTGAAATCAGGGAACAACCTGCTATGGCTGCACACCTCCTTCGCCTTCCTGTAT CTGCTGCTCACCGTCTACAGCATGCGTAGACACACCTCCAAGATGCGCTACAAGGAGGATGATCTGG TGCGTCGGACCCTCTTCATCAATGGAATCTCCAAATATGCAGAGTCAGAAAAGATCAAGAAGCATTT TAGGGAAGCCTACCCCAACTGCACAGTTCTCGAAGCCCGCCCGTGTTACAACGTGGCTCGCCTAATG TTCCTCGATGCAGAGAGGAAGAAGGCCGAGCGGGAAAGCTGTACTTCACAAACCTCCAGAGCAAGG AGAACGTGCCTACCATGATCAACCCCAAGCCCTGTGGCCACCTCTGCTGCTGTGTGGTGCGAGGCTG TGAGGAGGCCATTGAGTACTACACAAAGCTGGAGCAGAAGCTGAAGGAAGACTACAAGCGGGAGAAG GAGAAGGTGAATGAGAAGCCTCTTGGCATGGCCTTTGTCACCTTCCACAATGAGACTATCATCCTGA AGGACTTCAACGTGTGTAAATGCCAGGGCTGCACCTGCCGTGGGGAGCCACGCCCCTCATCCTGCAG CGAGTCCCTGCACATCTCCAACTGGACCGTGTCCTATGCCCCTGACCCTCAGAACATCTACTGGGAG CACCTCTCCATCCGAGGCTTCATCTGGTGGCTGCGCTGCCTGGTCATCAATGTCGTCCTCTTCATCC TCCTCTTCTTCCTCACCACTCCAGCCATCATCATCACCACCATGGACAAGTTCAACGTCACCAAGCC TGTGGAGTACCTCAACAACCCCATCATCACCCAGTTCTTCCCCACCCTGCTGCTGTGGTGCTTCTCG GCCCTCCTTCCCACCATCGTCTACTACTCAGCCTTCTTTGAAGCCCACTGGACACGGTCCAGCTCTG GGGAGAACAGGACAACCATGCACAAGTGCTACACTTTCCTCATCTTCATGGTGCTGCTCCTACCCTC GCTGGGACTGAGCAGCCTGGACCTCTTCTTCCGCTGGCTCTTTGATAAGAAATTCTTGGCTGAGGCA GCTATTCGGTTTGAGTGTGTGTTCCTGCCCGACAACGGCGCCTTCTTCGTGAACTACGTCATTGCCT CAGCCTTTATCGGCAACGCCATGGACCTGCTGCGCATCCCAGGCCTGCTCATGTACATGATCCGGCT CTGCCTGGCGCGCTCGGCCGCCGAGAGGCGCAACGTGAAGCAGCATCAGGCCTACGAGTTCCAGTTT GGCGCAGCCTACGCCTGGATGATGTGCGTCTTCACGGTGGTCATGACCTACAGTATCACCTGCCCCA TCATCGTGCCCTTCGGGCTCATGTACATGCTGCTGAAGCACCTGGTAGACAGGTACAATCTCTACTA CGCCTACCTGCCGGCCAAGCTGGACAAGAAGATCCACTCGGGGGCTGTGAACCAGGTGGTGGCCGCG CCCATCCTCTGCCTCTTCTGGCTGCTCTTCTTTTCCACCATGCGCACGGGGTTCCTAGCTCCCACGT CTATGTTCACATTTGTGGTCCTGGTCATCACCATCGTCATCTGTCTCTGCCACGTCTGCTTTGGACA CTTCAAATACCTCAGTGCCCACAACTACAAGATTGAGCACACGGAGACAGATACTGTGGACCCCAGA AGCAATGGACGGCCCCCCACTGCTGCTGCTGTCCCCAAATCTGCGAAATACATCGCTCAGGTGCTGC AGGACTCAGAGGTGGACGGGGATGGGGATGGGGCTCCTGGGAGCTCAGGGGATGAGCCCCCATCATC CTCATCCCAAGATGAGGAGTTGCTGATGCCACCCGACGCCCTCACGGACACAGACTTCCAGTCTTGC GAGGACAGCCTCATAGAGAATGAGATTCACCAG**TAA**GGGGAGGGAGGGGCCCTGGAGGCCACATCCT GCCCCACCCCACCCCCACTCCCACGGACACTAAAACGCTAATAATTTATTAGATCTAAAGCCCCTTC CTCCCCAGCCCCTGCTTTCATTAAGGTATTTAAACTTGGGGGTTTCACTGCTCTCCCCCCATGATGG AGGGAGGGAGCCCCCCAACCTCAGTGAGGAGAGCCCAGAGCCGGCCCCGGGGCAAAGAGGGGTGCAG AGGGAGTTCCCCCAGATCAGTACCCCCAACACCTCACCAGATAGTAGCAAGCACCAAAAGAGGGTTA ATGAGAGCCAAGAGGAGTACCTGGTGCACCTGGTGCCGGTGGCTGGAGACCTGGGGGGCAGGTGGAT CTGGGGCTGTTCCCCCCCCTCCGTTTTTTCCACCCCACAGTTCCTCCTGGGATCTGGCCCTCCAGGG AAGTGGAGCCTCCAGCCCCTAGGGGATGCATGAGGGGGGAGGGGGTGCTGAGTGGGAGGAAGAGTCA GGCTCACAGCTGGGGTGGCCTGGGGGTGGGGGTGGGCAAGGCTGACACTGGAAAATGGGTTTTTGCA CTGTTTTTTTTTTTGGTTTTTTTTGTTCTTTTTTTGTTTTTTTCCTTTAAAATAAAAACAAAGAAAAGCT CTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAG		
	ORF Start: ATG at 16		ORF Stop: TAA at 2446

	SEQ ID NO: 86	810 aa	MW at 92305.1kD
NOV20a, CG145978-01 Protein Sequence	MLPFLLATLGTTALNNSNPKDYCYSARIRSTVLQGLPFGGVPTVLALDFMCFLFPQALLFLFSILRK VAWDYGRLALVTDADSHDRYERLTSVSSSVDFDQRDNVGFCSWLTAIFRIDDEIRDKCGGDAVHYLS FQRHIIGLLVVVGVLSVGIVLPVNFSGDLLENNAYSFGRTTIANLKSGNNLLWLHTSFAFLYLLLTV YSMRRHTSKMRYKEDDLVRRTLFINGISKYAESEKIKKHFREAYPNCTVLEARPCYNVARLMFLDAE RKKAERGKLYFTNLQSKENVPTMINPKPCGHLCCCVVRGCEEAIEYYTKLEQKLKEDYKREKEKVNE KPLGMAFVTFHNETIILKDFNVCKCQGCTCRGEPRPSSCSESLHISNWTVSYAPDPQNIYWEHLSIR		

	GFIWWLRCLVINVVLFILLFFLTTPAIIITTMDKFNVTKPVEYLNNPIITQFFPTLLLWCFSALLPT IVYYSAFFEAHWTRSSSGENRTTMHKCYTFLIFMVLLLPSLGLSSLDLFFRWLFDKKFLAEAAIRFE CVFLPDNGAFFVNYVIASAFIGNAMDLLRIPGLLMYMIRLCLARSAAERRNVKQHQAYEFQFGAAYA WMMCVFTVVMTYSITCPIIVPFGLMYMLLKHLVDRYNLYYAYLPAKLDKKIHSGAVNQVVAAPILCL FWLLFFSTMRTGFLAPTSMFTFVVLVITIVICLCHVCFGHFKYLSAHNYKIEHTETDTVDPRSNGRP PTAAAVPKSAKYIAQVLQDSEVDGDGDGAPGSSGDEPPSSSSQDEELLMPPDALTDTDFQSCEDSLI ENEIHQ

	SEQ ID NO: 87	1864 bp	
NOV20b, CG145978-02 DNA Sequence	GCCGCCCAGCGACTCCCCCTCCCCCTCCCCCAGCCCCGCCCCGCCCCAACCCGGGGCTCCGAGCCGG AGCCGAGTCTGCGCCTGGGGAGGACCATGCGGCAGTAGCAGCC**ATG**CTGCCCTTTCTGCTGGCCAC ACTGGGCACCACAGCCCTCAACAACAGCAACCCCAAGGACTACTGCTACAGCGCCCGCATCCGCAGC ACTGTCCTGCAGGGCCTGCCCTTTGGGGGCGTCCCCACCGTGCTGGCTCTCGACTTCATGTGCTTCC TTCCTCAGGCACTGCTGTTCTTATTCTCTATCCTCCGGAAGGTGGCCTGGGACTATGGGCGGCTGGC CTTGGTGACAGATGCAGACAGGCTTCGGCGGCAGGAGAGGGACCGAGTGGAACAGGAATATGTGGCT TCAGCTATGCACGGGGACAGCCATGACCGGTATGAGCGTCTCACCTCTGTCTCCAGCTCCGTTGACT TTGACCAAAGGGACAATGTGGGTTTCTGTTCCTGGCTGACAGCCATCTTCAGGATAAAGGATGATGA GATCCGGGACAAATGTGGGGGCGACGCCGTGCACTACCTGTCCTTTCAGCGGCACATCATCGGGCTG CTGGTGGTTGTGGGCGTCCTCTCCGTAGGCATCGTGCTGCCTGTCAACTTCTCAGGGGACCTGCTGG AGAACAATGCCTACAGCTTTGGGAGAACCACCATTGCCAACTTGAAATCAGGGAACAACCTGCTATG GCTGCACACCTCCTTCGCCTTCCTGTATCTGCTGCTCACCGTCTACAGCATGCGTAGACACACCTCC AAGATGCGCTACAAGGAGGATGATCTGGTGCGTCGGACCCTCTTCATCAATGGAATCTCCAAATATG CAGAGTCAGAAAAGATCAAGAAGCATTTTAGGGAAGCCTACCCCAACTGCACAGTTCTCGAAGCCCG CCCGTGTTACAACGTGGCTCGCCTAATGTTCCTCGATGCAGAGAGGAAGAAGGCCGAGCGGGGAAAG CTGTACTTCACAAACCTCCAGAGCAAGGAGAACGTGCCTACCATGATCAACCCCAAGCCCTGTGGCC ACCTCTGCTGCTGTGTGGTGCGAGGCTGTGAGCAGGTGGAGGCCATTGAGTACTACACAAAGCTGGA GCAGAAGCTGAAGGAAGACTACAAGCGGGAGAAGGAGAAGGTGAATGAGAAGCCTCTTGGCATGGCC TTTGTCACCTTCCACAATGAGACTATCACCGCCATCATCCTGAAGGACTTCAACGTGTGTAAATGCC AGGGCTGCACCTGCCGTGGGGAGCCACGCCCCTCATCCTGCAGCGAGTCCCTGCACATCTCCAACTG GACCGTGTCCTATGCCCCTGACCCTCAGAACATCTACTGGGAGCACCTCTCCATCCGAGGCTTCATC TGGTGGCTGCGCTGCCTGGTCATCAATGTCGTCCTCTTCATCCTCCTCTTCTTCCTCACCACTCCAG CCATCATCATCACCACCATGGACAAGTTCAACGTCACCAAGCCTGTGGAGTACCTCAACGTGAGGCC TCATGCCCCTGTCACTTTCCACGCTGGGTCACAACACACAGATACCAGGCCC**TGA**TCCCTCTTCCAC TTGCCCAGCCCAGCCCGTTCTGCTTGTTCCAACCCCGTGCCACCAACCAGCTCCCAAAAACCCCTGT GTGCACTTCCCTTGGGCTCCCTGCCACCTTCCCCCTGAGAGAGGCCACCCTCAGGTGTGCAACACCT GGAGAAACACCCAGGTAAGAGAGAGAGCCTGCATTTAGTCCTGATCTCAGAGAAGTCCCCTTCCCTC ACCCCTCAGTCTAACTGAAAAAATGGAAAGGTTTGACTAGAAAAAAAAAAAAAAA		
	ORF Start: ATG at 112		ORF Stop: TGA at 1594

	SEQ ID NO: 88	494 aa	MW at 56686.9kD
NOV20b, CG145978-02 Protein Sequence	MLPFLLATLGTTALNNSNPKDYCYSARIRSTVLQGLPFGGVPTVLALDFMCFLPQALLFLFSILRKV AWDYGRLALVTDADRLRRQERDRVEQEYVASAMHGDSHDRYERLTSVSSSVDFDQRDNVGFCSWLTA IFRIKDDEIRDKCGGDAVHYLSFQRHIIGLLVVVGVLSVGIVLPVNFSGDLLENNAYSFGRTTIANL KSGNNLLWLHTSFAFLYLLLTVYSMRRHTSKMRYKEDDLVRRTLFINGISKYAESEKIKKHFREAYP NCTVLEARPCYNVARLMFLDAERKKAERGKLYFTNLQSKENVPTMINPKPCGHLCCCVVRGCEQVEA IEYYTKLEQKLKEDYKREKEKVNEKPLGMAFVTFHNETITAIILKDFNVCKCQGCTCRGEPRPSSCS ESLHISNWTVSYAPDPQNIYWEHLSIRGFIWWLRCLVINVVLFILLFFLTTPAIIITTMDKFNVTKP VEYLNVRPHAPVTFHAGSQHTDTRP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 20B.

Table 20B. Comparison of NOV20a against NOV20b.		
Protein Sequence	**NOV20a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV20b	1..447 1..474	394/475 (82%) 394/475 (82%)

Further analysis of the NOV20a protein yielded the following properties shown in Table 20C.

Table 20C. Protein Sequence Properties NOV20a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 14 and 15

A search of the NOV20a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 20D.

Table 20D. Geneseq Results for NOV20a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB97369	Novel human protein SEQ ID NO: 637 - Homo sapiens, 541 aa. [WO200222660-A2, 21-MAR-2002]	291..792 1..505	499/507 (98%) 500/507 (98%)	0.0
AAB94004	Human protein sequence SEQ ID NO:14117 - Homo sapiens, 807 aa. [EP1074617-A2, 07-FEB-2001]	19..745 27..755	445/735 (60%) 565/735 (76%)	0.0
AAB42245	Human ORFX ORF2009 polypeptide sequence SEQ ID NO:4018 - Homo sapiens, 480 aa. [WO200058473-A2, 05-OCT-2000]	19..472 3..480	440/482 (91%) 442/482 (91%)	0.0
ABG63456	Human albumin fusion protein #131 - Homo sapiens, 318 aa. [WO200177137-A1, 18-OCT-2001]	493..810 1..318	316/318 (99%) 318/318 (99%)	0.0
AAG71250	Human gene 8-encoded secreted protein HCEIE80, SEQ ID NO:98 - Homo sapiens, 318 aa. [WO200132674-A1, 10-MAY-2001]	493..810 1..318	316/318 (99%) 318/318 (99%)	0.0

In a BLAST search of public sequence datbases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20E.

Table 20E. Public BLASTP Results for NOV20a				
Protein Accession Number	**Protein/Organism/Length**	**NOV20a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
CAD38916	Hypothetical protein - Homo sapiens (Human), 519 aa (fragment).	167..681 1..518	510/520 (98%) 512/520 (98%)	0.0
AAH30245	KIAA0792 gene product - Homo sapiens (Human), 807 aa.	19..745 27..755	449/735 (61%) 570/735 (77%)	0.0
O94886	KIAA0792 protein - Homo sapiens (Human), 807 aa.	19..745 27..755	448/735 (60%) 569/735 (76%)	0.0
Q91YT8	Hypothetical 91.9 kDa protein - Mus musculus (Mouse), 804 aa.	19..745 27..754	446/735 (60%) 570/735 (76%)	0.0
BAC04207	CDNA FLJ36310 fis, clone THYMU2005001 - Homo sapiens (Human), 491 aa.	1..447 1..471	440/475 (92%) 441/475 (92%)	0.0

PFam analysis predicts that the NOV20a protein contains the domains shown in the Table 20F.

Table 20F. Domain Analysis of NOV20a			
Pfam Domain	**NOV20a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
DUF221	327..787	109/493 (22%) 365/493 (74%)	1.1e-84

Example 21.

The NOV21 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis			
	SEQ ID NO: 89	8700 bp	
NOV21a, CG145997-01	GCTGATTGAGACACTATGTTGAGTCTACAGGATTCTGTGTTTTTGAAATTAGCATAAAGTCCTTGT TAAAGTCCTGGAGCAGCAGCATGTCAAACATTACGATTGACCCAGATGTCAAACCTGGTGAATATGT		

DNA Sequence	CATCAAGAGCCTCTTTGCAGAATTTGCTGTTCAAGCTGAAAAGAAAATTGAAGTTGTAATGGCCGAA CCCTTGGAGAAGCTATTGTCCAGATCTCTTCAGAGGGGTGAAGATCTTCAGTTTGATCAGGTAATAA GCTCTATGAGCTCAGTAGCAGAGCACTGTCTCCCTTCCTTACTTCGCACCTTGTTGACTGGTACAG ACGCCAAAATGGAACGGAAGATGAATCTTATGAATATAGGCCTCGGTCTAGCACAAAGTCTAAGGAT GAACAGCAACGTGAAAGAGATTATCTTCTTGAAAGGAGGGACTTAGCAGTAGACTTCATTTTTTGTT TAGTTTTAGTTGAAGTTCTAAAGCAGATTCCTGTTCATCCTGTACCCGATCCCTTAGTTCATGAAGT TCTAAACTTAGCTTTTAAGCACTTTAAACATAAGGAAGGGGGAACCAACACTGGGAATGTGCATATT ATTGCTGATTTATATGCAGAGGTGATAGGGGTTCTTGCCCAATCAAAGTTTCAGGCTGTAAGGAAGA AGTTTGTGACAGAATTAAAAGAACTGCGACAAAAGGAACAAAGCCCACATGTGGTACAAAGTGTCAT CAGCTTAATAATGGGAATGAAATTTTTTCGAGTAAAAATGTATCCTGTAGAAGATTTTGAAGCATCA TTTCAATTTATGCAGGAATGTGCTCAGTATTTCTTAGAAGTGAAAGATAAAGATATAAAACATGCAC TTGCTGGTTTATTTGTGGAGATTCTTATCCCTGTAGTTAAAAATGAAGTGAATGTTCCCTGTTTGAA AAATTTTGTGGAGATGCTTTATCAGACTACTTTTGAACTGAGCTCGAGAAAGAAGCATTCATTGGTA TTAAATAAAGATCCGAAAATGTCTCGAGTTGCACTGGAATCTTTGTATAGATTATTGTGGGTTTATG TAATTAGAATAAAATGTGAAAGCAACACTGTAACTCAAAGTCGTCTTATGAGCATAGTGTCAGCACT TTTTCCAAAAGGCTCACGAAGTGTGGTTCCTCGTGACACACCTCTCAATATATTTGTGAAGATTATT CAGTTCATTGCTCAGGAACGCTTGGATTTTGCAATGAAAGAAATAATATTTGATCTTCTCAGTGTTG GAAAATCTACTAAAACTTTCACCATTAATCCAGAGTGTCTAGCATATGTAATATGTTTCTTATTAAA TCCTGTTGTATTTTTCACGGGGGAAAGAAAACCCAAGATTGATTTGTTTAGAACTTGTATTGCTGCG ATTCCAAGGTTGATTCCTGACGGTATGAGCAGAACTGACCTGATTGAATTGTTAGCAAGGCTCACAA TTCATATGGATGAAGAACTGCGTGCTCTGGCTTTCAATACTCTGCAGGCACTAATGCTTGATTTTCC AGATTGGCGGGAGGATGTTCTTTCAGGATTTGTTTATTTTATTGTTCGTGAAGTGACTGATGTCCAT CCCACACTTCTTGATAATGCCGTAAAGATGTTGGTACAATTAATAAATCAGTGGAAACAAGCAGCCC AAATGCATAATAAAAACCAGGACACTCAGGTACCAGATTCTTTTCTAGTAGCTAATGGAGCTTCTCA TCCCCCTCCTCTGGAAAGGAGCCCATATTCCAATGTATTCCATGTGGTTGAAGGCTTTGCGCTTGTC ATTCTCTGTAGCAGTCGACCTGCCACTAGGAGACTAGCCGTCAGTGTCCTTAGAGAAATACGGGCTT TATTTGCACTTCTGGAAATACCTAAGGGTGATGAGTGAATTAGCCATAGATGTGATGGACAGGCTAAG CCCATCCATTCTTGAGAGTTTCATACATCTCACTGGGGCTGATCAGGTAACTATATCGATAGATTTA CAAACTTTAGCAGAATGGAACTCTTCTCCTATTAGCCACCAGTTTGATGTGATTAGTCCATCACATA TATGGATATTTGCACATGTGACCCAAGGCCAAGACCCATGGATTATAAGTCTCTCCAGTTTTTTAAA GCAAGAAAATCTTCCTAAACACTGCTCTACAGCTGTGAGCTATGCTTGGATGTTTGCATACACAAGA CTTCAGTTGTTGTCCCCTCAGGTCGATAGTAGCCCCATCAATGCTAAGAAAGTAAATACCACCACAA GCAGTGACTCATACATTGGCCTGTGGAGAAACTATCTGATCCTTTGCTGCAGTGCAGCAACATCGTC ATCTTCCACATCTGCAGGTTCTGTGAGATGTTCTCCTCCTGAGACGCTGGCGTCTACCCCAGATAGC GGCTATAGCATTGATTCTAAAATTGGCATCCCATCCCCTTCATCCTTGTTTAAGCACATAGTTCCAA TGATGCGTTCTGAGAGCATGGAAATCACAGAATCCCTTGTTCTAGGTCTTGGCAGGACCAACCCAGG AGCTTTTAGGAATATGAAACGGCGCAGGCGTCGAGACATTTTACGAGTACAACTGGTACGAATATTT GAACTGCTGGCAGATGCTGGTGTCATTAGTAGTGCAAGTGGTGGCCTTGATAATGAAACACATTTTC TCAACAACACTTTATTGGAATATGTAGATTTAACTAGACAACTCCTGGAAGCAGAAAATGAAAAAGA CTCTGACACACTGAAGGATATACGATGCCATTTTAGTGCCTTAGTGGCGAATATTATTCAGAATGTT CCAGTGCACCAGAGAAGAAGTATTTTTCCTCAACAGAGCCTTCGTCACAGTCTATTTATGCTGTTCA GTCACTGGGCAGGTCCTTTTAGCATCATGTTTACGCCCTTGGACAGATACAGTGATAGAAATATGCA AATTAATAGACATCAATACTGTGCGTTAAAGGCTATGTCTGCTGTACTGTGTTGTGGCCCTGTTGCA GATAATGTAGGACTTTCATCAGATGGCTATTTGTACAAATGGTTGGATAACATTTTGGATTCTCTGG ACAAAAAGGTTCACCAGCTGGGCTGTGAAGCAGTTACGTTGTTACTGGAGCTGAACCCTGATCAGAA CAACCTGATGTACTGGGCCAGGGATTATCAATGTGACACAGTGATGCTTCTAAATCTGATACTGTTT AAAGCAGCTGATTCTTCTAGAAGTATCTATGAAGTTGCTATGCAACTTTTACAGATTCTGGAACCGA AGATGTTTCGCTATGCTCACAAATTGGAGGTTCAGAGAACAGATGGAGTACTCAGCCAGCTGTCTCC TCTACCACATCTCTATTCTGTTTCATATTATCAGTTGTCCGAGGAACTAGCAAGGGCGTATCCTGAG CTAACTCTCGCCATATTCTCAGGTAAGCCAGAGAATCCAGACAGCTCACCCTGCTGGGCGGCAGGTG ATGCTGCACTACCTGCTACCATGGATGAACAACATCGAGCTGGTGGACTGCAAGCTCCGCCTCCCGG GTTCACGCCATTCTCCGATGATTCCTTAAAAGACCGAGAACTTATGGTGACTAGTAGGCGCTGGTTA CGGGGAGAAGGATGGGGATCTCCACAAGCCACTGCAATGGTTTTGAACAATCTGATGTATATGACAG CAAAGTATGGCGATGAACTGGCCTGGTCGGAGGTGGAGAATGTGTGGACCACACTTGCAGATGGCTG GCCCAAAAACCTGAAAATAATTTTGCACTTTTTGATCAGCATTTGTGGGGTGAATAGCGAACCAAGC CTCTTGCCTTACGTAAAGAAGGTCATTGTATATTTAGGTAGAGATAAAACAATGCAGTTGCTAGAAG AGCTGGTGAGTGAGCTTCAGCTGACCGATCCTGTCAGTTCAGGGGTCACTCACATGGATAATCCCCC GTATTATCGCATCACTTCCAGCGCTTTGTCTTTGATTACAGGAACTACTTCCAGTAGCAATACAATG GTAGCTCCCACAGATGGCAATCCTGATAATAAGCCCATTAAAGAGAATATTGAAGAGAGGACCAGTC ATTTGAATCGGCAACATCCCAGCCTAGAATCCCGATACAGTAGCAGCTCTGGAGGATCTTATGAAGA AGAAAAAAGTGATTCAATGCCACTTTATTCTAATTGGCGACTGAAAGTGATGGAGCATAACCAAGGA GAGCCACTGCCCTTCCCACCAGCTGGAGGCTGCTGGTCACCACTGGTGGATTACGTGCCTGAAACGT CATCACCTGGATTACCTCTTCACAGGTGTAACATAGCAGTGATCCTTTTGACTGATCTCATCATTGA TCATAGTGTGAAGGTGGAATGGGGAAGCTACCTCCATCTTCTTCTTCATGCAATTTTTTAGGGTTT GACCACTGCCACCCTGAGGTGTATGAACATTGTAAACGCCTGCTTCTGCACTTATTAATAGTAATGG GACCCAATAGTAACATCCGAACTGTTGCTTCTGTCCTTCTCAGGAACAAGGAGTTTAATGAGCCCAG GGTGCTTACAGTCAAACAAGTTGCACACTTAGATTATAATTTCACAGGTATTAACGATTTTATACCT GATTACCAGCCCTCCCCTATGACTGACTCAGGGCTTAGCTCAAGTTCTACCTCTTCTAGTATCAGCT TAGGAAATAACAGTGCTGCCATTTCACATCTGCACACCACTATCCTCAATGAGGTTGACATCTCAGT GGAGCAGGATGGAAAAGTCAAAACCCTCATGGAATTCATTACCTCAAGGAAAAGAGGGCCCCTTTGG AACCATGAGGATGTTTCTGCCAAGAATCCTAGCATAAAGAGTGCTGAACAGTTAACTACATTTTTGA AACATGTGGTTTCTGTTTTTAAGCAGTCAAGCTCAGAAGGAATTCATCTGGAACATCATCTTAGTGA AGTTGCTCTGCAAACAGCACTTTCCTGTTCTTCTCGACACTATGCTGGGAGATCCTTTCAGATTTTC AGGGCCCTAAAGCAGCCTCTCACTGCAACTACACTTTCTGATGTTCTCTCCAGACTTGTAGAAACTG TAGGGGATCCAGGAGAAGATGCACAGGGATTTGTGATTGAGCTTCTTCTCACATTGGAATCTGCAAT TGATACTTTGGCTGAAACCATGAAGCATTATGATCTTCTTTCTGCCCTTTCTCAAACCTCATATCAT

	GATCCTATAATGGGAAACAAGTATGCAGCTAACAGGAAAAGCACTGGACAACTCAATCTAAGCACAA GTCCCATTAATAGTAGCAGTTATTTGGGATATAACAGTAATGCAAGAAGTAACTCTTTGAGATTAAG TTTGATTGGTGACCGACGAGGTGACCGGCGGCGGAGTAACACACTGGATATAATGGATGGACGGATA AACCATAGCAGTAGTTTAGCAAGGACTAGAAGCCTTTCCTCTCTAAGAGAGAAAGGAATGTATGACG TGCAGTCCACTACTGAGCCTACCAACTTGATGGCCACCATTTTTTGGATAGCAGCATCTTTATTAGA ATCAGATTATGAATATGAATACCTCCTGGCTCTCAGGCTTCTCAACAAACTGCTTATCCATTTGCCT TTGGATAAATCAGAGAGTCGAGAGAAGATTGAAAATGTACAAAGCAAATTGAAATGGACTAATTTTC CAGGACTTCAGCAGCTCTTCCTTAAGGGTTTACCTCAGCATCTACACAAGAAATGACCGTGCACCT CCTCAGTAAACTCATTTCTGTCTCCAAACATACATTGGTGGATCCTTCCCAATTGTCAGGCTTTCCT CTTAACATCCTTTGCTTATTGCCTCACTTAATCCAGCATTTTGACAGCCCAACTCAGTTTTGCAAAG AAACAGCTAGTCGAATAGCAAAGGTTTGTGCAGAAGAAAAATGCCCAACACTTGTCAATCTGGCACA CATGATGAGTTTGTACAGTACACACACGTATTCCAGAGACTGTTCTAACTGGATCAATGTCGTGTGC AGATACCTGCATGACTCCTTCTCAGATACAACATTTAATCTTGTGACTTATCTTGCAGAGCTGTTAG AGAAAGGATTGTCCAGTATGCAGCAATCATTACTACAGATTATTTATAGTCTATTGAGTCATATTGA CCTGTCTGCAGCCCCAGCCAAGCAGTTTAATCTGGAGATCATAAAGATTATTGGCAAATATGTACAG AGTCCTTACTGGAAGGAAGCCCTTAACATATTAAAGCTGGTGGTGTCACGCTCTGCGAGTCTTGTCG TACCCAGTGATATCCCCAAGACCTATGGAGGAGATACAGGTTCTCCTGAAATATCCTTCACTAAAAT TTTTAATAATGTTTCTAAGGAGTTGCCTGGGAAGACCTTAGATTTTCATTTTGATATATCTGAGACA CCAATTATTGGAAACAAATATGGTGATCAGCACAGTGCGGCTGGAAGAAATGGGAAACCAAAAGTTA TTGCTGTCACTAGAAGTACTTCCTCAACTTCTTCTGGTTCTAATTCTAATGCCTTGGTTCCTGTTAG TTGGAAAAGGCCACAGTTATCACAGCGAAGAACAAGAGAAAAGCTAATGAATGTGCTTTCTCTCTGT GGTCCAGAATCTGGCCTCCCAAAGAACCCATCAGTTGTATTTTCTTCTAATGAGGATTTGGAAGTCG GTGACCAACAGACTAGCCTAATTTCTACAACAGAAGACATAAATCAAGAGGAAGAAGTAGCTGTGGA AGATAATAGCAGTGAACAACAGTTTGGTGTTTTTAAGGATTTTGACTTTTTAGATGTTGAATTGGAA GATGCAGAGGGTGAAAGTATGGACAATTTCAACTGGGGAGTTCGCAGGCGCTCACTGGACAGTATTG ACAAAGGGGACACTCCATCCCTCCAGGAGTACCAGTGCTCTAGTAGCACCCCCAGCCTGAACCTCAC CAATCAGGAGGATACAGATGAGTCCTCGGAAGAAGAAGCGGCACTTACAGCAAGCCAGATACTCTCA CGCACACAGATGTTAAACAGTGATTCTGCCACTGATGAAACAATACCAGACCATCCTGACTTACTTC TCCAGTCTGAAGATTCCACTGGCAGCATCACAACAGAGGAAGTGCTTCAAATCAGGGATGAGACCCC AACTTTGGAGGCTTCTCTAGATAATGCTAACAGCCGGCTGCCTGAGGATACAACTTCAGTATTAAAG GAGGAACATGTTACAACCTTTGAAGATGAAGGATCCTATATAATTCAAGAACAGCAGGAATCTCTTG TGTGTCAAGGAATTCTTGATTTAGAAGAAACTGAAATGCCAGAGCCTCTAGCTCCTGAAAGTTACCC CGAGTCAGTCTGTGAAGAGGATGTTACCTTAGCTCTGAAAGAGCTAGATGAAAGATGTGAAGAAGAA GAAGCGGATTTCTCCGGACTGTCTAGTCAAGATGAAGAAGAGCAAGATGGTTTTCCAGAAGTACAGA CGTCGCCTCTGCCGTCACCATTTCTTTCTGCCATCATAGCCGCCTTTCAGCCCGTGGCATATGATGA TGAAGAGGAAGCCTGGCGCTGCCACGTCAATCAGATGCTGTCTGACACCGACGGGTCCTCTGCAGTG TTTACTTTTCATGTGTTTTCTAGGCTGTTTCAGACAATTCAAAGAAAGTTTGGAGAAATAACTAATG AGGCAGTCAGCTTTCTTGGTGATAGTCTGCAACGCATTGGTACCAAATTTAAAAGTTCCTTGGAAGT GATGATGCTGTGTTCAGAATGCCCAACAGTCTTTGTGGATGCTGAAACACTGATGTCATGTGGTTTG CTGGAAACACTCAAGTTTGGTGTTTTGGAGTTGCAAGAACACCTGGATACATACAATGTGAAAAGAG AAGCCGCTGAGCAGGAATTGGAGCTCTGCCGAAGATTATACAAATTGCATTTTCAATTGCTGCTTCT GTTCCAGGCCTACTGTAAACTTATCAACCAAGTAAATACGATAAAAAATGAAGCAGAGGTCATCAAC ATGTCAGAGGAACTTGCCCAACTGGAAAGTATCCTCAAAGAAGCTGAGTCCGCTTCCGAAAACGAAG AAATTGACATTTCCAAAGCTGCACAAACTACTATAGAAACTGCCATTCATTCTTTAATTGAAACTTT GAAAAATAAAGAATTTATATCAGCTGTAGCACAAGTCAAAGCTTTCAGATCTCTCTGGCCCAGTGAT ATCTTTGGCAGTTGTGAAGATGACCCTGTACAGACACTGTTACATATATATTTCCATCATCAGACGC TGGGCCAGACAGGAAGCTTTGCAGTTATAGGCTCTAACCTGGACATGTCAGAAGCCAACTACAAACT GATGGAACTTAATCTGGAAATAAGAGAGTCTCTACGCATGGTGCAATCATACCAACTTCTAGCACAG GCCAAACCAATGGGAAATATGGTGAGCACTGGATTCTGAGACACTTCAGGCCTTTAGGAAAGAAACT AAACTGAAGATGATGAAGAATATTAACCAAGCACCTTTTATGGACCCTTGCATTCACTGATAACTTT CTGGCAGCATCTACTTTTTAGTGTAACTAATGTCAAACTGTATCATCAAAAACAAAGATCTGAAAGA AAAAAACATCTGATATTTTAACAGCTGCCAATATCTCCCACAATAACTGCGTGAAGA		
	ORF Start: ATG at 16		ORF Stop: TGA at 8479

	SEQ ID NO: 90	2821 aa	MW at 316987.5kD
NOV21a, CG145997-01 Protein Sequence	MLSLQDSVFFEISIKSLLKSWSSSMSNITIDPDVKPGEYVIKSLFAEFAVQAEKKIEVVMAEPLEKL LSRSLQRGEDLQFDQVISSMSSVAEHCLPSLLRTLFDWYRRQNGTEDESYEYRPRSSTKSKDEQQRE RDYLLERRDLAVDFIFCLVLVEVLKQIPVHPVPDPLVHEVLNLAFKHFKHKEGGTNTGNVHIIADLY AEVIGVLAQSKFQAVRKKFVTELKELRQKEQSPHVVQSVISLIMGMKFFRVKMYPVEDFEASFQFMQ ECAQYFLEVKDKDIKHALAGLFVEILIPVVKNEVNVPCLKNFVEMLYQTTFELSSRKKHSLVLNKDP KMSRVALESLYRLLWVYVIRIKCESNTVTQSRLMSIVSALFPKGSRSVVPRDTPLNIFVKIIQFIAQ ERLDFAMKEIIFDLLSVGKSTKTFTINPECLAYVICFLLNPVVFFTGERKPKIDLFRTCIAAIPRLI PDGMSRTDLIELLARLTIHMDEELRALAFNTLQALMLDFPDWREDVLSGFVYFIVREVTDVHPTLLD NAVKMLVQLINQWKQAAQMHNKNQDTQVPDSFLVANGASHPPPLERSPYSNVFHVVEGFALVILCSS RPATRRLAVSVLREIRALFALLEIPKGDDELAIDVMDRLSPSILESFIHLTGADQVTISIDLQTLAE WNSSPISHQFDVISPSHIWIFAHVTQGQDPWIISLSSFLKQENLPKHCSTAVSYAWMFAYTRLQLLS PQVDSSPINAKKVNTTTSSDSYIGLWRNYLILCCSAATSSSSTSAGSVRCSPPETLASTPDSGYSID SKIGIPSPSSLFKHIVPMMRSESMEITESLVLGLGRTNPGAFRNMKRRRRRDILRVQLVRIFELLAD AGVISSASGGLDNETHFLNNTLLEYVDLTRQLLEAENEKDSDTLKDIRCHFSALVANIIQNVPVHQR RSIFPQQSLRHSLFMLFSHWAGPFSIMFTPLDRYSDRNMQINRHQYCALKAMSAVLCCGPVADNVGL SSDGYLYKWLDNILDSLDKKVHQLGCEAVTLLLELNPDQNNLMYWARDYQCDTVMLLNLILFKAADS SRSIYEVAMQLLQILEPKMFRYAHKLEVQRTDGVLSQLSPLPHLYSVSYYQLSEELARAYPELTLAI		

	FSGKPENPDSSPCWAAGDAALPATMDEQHRAGGLQAPPPGFTPFSDDSLKDRELMVTSRRWLRGEGW GSPQATAMVLNNLMYMTAKYGDELAWSEVENVWTTLADGWPKNLKIILHFLISICGVNSEPSLLPYV KKVIVYLGRDKTMQLLEELVSELQLTDPVSSGVTHMDNPPYYRITSSALSLITGTTSSSNTMVAPTD GNPDNKPIKENIEERTSHLNRQHPSLESRYSSSSGGSYEEEKSDSMPLYSNWRLKVMEHNQGEPLPF PPAGGCWSPLVDYVPETSSPGLPLHRCNIAVILLTDLIIDHSVKVEWGSYLHLLLHAIFLGFDHCHP EVYEHCKRLLLHLLIVMGPNSNIRTVASVLLRNKEFNEPRVLTVKQVAHLDYNFTGINDFIPDYQPS PMTDSGLSSSSTSSSISLGNNSAAISHLHTTILNEVDISVEQDGKVKTLMEFITSRKRGPLWNHEDV SAKNPSIKSAEQLTTFLKHVVSVFKQSSSEGIHLEHHLSEVALQTALSCSSRHYAGRSFQIFRALKQ PLTATTLSDVLSRLVETVGDPGEDAQGFVIELLLTLESAIDTLAETMKHYDLLSALSQTSYHDPIMG NKYAANRKSTGQLNLSTSPINSSSYLGYNSNARSNSLRLSLIGDRRGDRRRSNTLDIMDGRINHSSS LARTRSLSSLREKGMYDVQSTTEPTNLMATIFWIAASLLESDYEYEYLLALRLLNKLLIHLPLDKSE SREKIENVQSKLKWTNFPGLQQLFLKGFTSASTQEMTVHLLSKLISVSKHTLVDPSQLSGFPLNILC LLPHLIQHFDSPTQFCKETASRIAKVCAEEKCPTLVNLAHMMSLYSTHTYSRDCSNWINVVCRYLHD SFSDTTFNLVTYLAELLEKGLSSMQQSLLQIIYSLLSHIDLSAAPAKQFNLEIIKIIGKYVQSPYWK EALNILKLVVSRSASLVVPSDIPKTYGGDTGSPEISFTKIFNNVSKELPGKTLDFHFDISETPIIGN KYGDQHSAAGRNGKPKVIAVTRSTSSTSSGSNSNALVPVSWKRPQLSQRRTREKLMNVLSLCGPESG LPKNPSVVFSSNEDLEVGDQQTSLISTTEDINQEEEVAVEDNSSEQQFGVFKDFDFLDVELEDAEGE SMDNFNWGVRRRSLDSIDKGDTPSLQEYQCSSSTPSLNLTNQEDTDESSEEEAALTASQILSRTQML NSDSATDETIPDHPDLLLQSEDSTGSITTEEVLQIRDETPTLEASLDNANSRLPEDTTSVLKEEHVT TFEDEGSYIIQEQQESLVCQGILDLEETEMPEPLAPESYPESVCEEDVTLALKELDERCEEEEADFS GLSSQDEEEQDGFPEVQTSPLPSPFLSAIIAAFQPVAYDDEEEAWRCHVNQMLSDTDGSSAVFTFHV FSRLFQTIQRKFGEITNEAVSFLGDSLQRIGTKFKSSLEVMMLCSECPTVFVDAETLMSCGLLETLK FGVLELQEHLDTYNVKREAAEQELELCRRLYKLHFQLLLLFQAYCKLINQVNTIKNEAEVINMSEEL AQLESILKEAESASENEEIDISKAAQTTIETAIHSLIETLKNKEFISAVAQVKAFRSLWPSDIFGSC EDDPVQTLLHIYFHHQTLGQTGSFAVIGSNLDMSEANYKLMELNLEIRESLRMVQSYQLLAQAKPMG NMVSTGF

Further analysis of the NOV21a protein yielded the following properties shown in Table 21B.

Table 21B. Protein Sequence Properties NOV21a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3538 probability located in mitochondrial inner membrane; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 21C.

Table 21C. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG04763	Novel human diagnostic protein #4754 - Homo sapiens, 3048 aa. [WO200175067-A2, 11-OCT-2001]	432..2819 609..3046	1382/2497 (55%) 1762/2497 (70%)	0.0
ABB97274	Novel human protein SEQ ID NO: 542 - Homo sapiens,	1591..2821 1..1254	1231/1254 (98%) 1231/1254 (98%)	0.0

	1254 aa. [WO200222660-A2, 21-MAR-2002]			
ABG04764	Novel human diagnostic protein #4755 - Homo sapiens, 2035 aa. [WO200175067-A2, 11-OCT-2001]	567..2276 293..1962	893/1794 (49%) 1155/1794 (63%)	0.0
AAB65130	Gene #26 associated peptide #21 - Homo sapiens, 703 aa. [WO200075375-A1, 14-DEC-2000]	2144..2821 3..703	675/701 (96%) 677/701 (96%)	0.0
AAB65110	Gene #26 associated peptide #1 - Homo sapiens, 702 aa. [WO200075375-A1, 14-DEC-2000]	2144..2820 3..702	675/700 (96%) 677/700 (96%)	0.0

In a BLAST search of public sequence datbases, the NOV21a protein was found to have homology to the proteins shown in the BLASTP data in Table 21D.

Table 21D. Public BLASTP Results for NOV21a				
Protein Accession Number	**Protein/Organism/Length**	**NOV21a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9Y3N6	Hypothetical 338.2 kDa protein - Homo sapiens (Human), 3012 aa.	432..2819 577..3010	1385/2493 (55%) 1764/2493 (70%)	0.0
O94915	KIAA0826 protein - Homo sapiens (Human), 1236 aa (fragment).	1615..2821 1..1236	1207/1236 (97%) 1207/1236 (97%)	0.0
O14572	WUGSC:H_248O15.1 protein - Homo sapiens (Human), 1849 aa (fragment).	449..2276 1..1849	1090/1892 (57%) 1375/1892 (72%)	0.0
Q91ZH1	DM505L19.1 (Novel protein) - Mus musculus (Mouse), 1595 aa (fragment).	1226..2819 1..1593	877/1652 (53%) 1152/1652 (69%)	0.0
O95640	Hypothetical 88.4 kDa protein - Homo sapiens (Human), 795 aa (fragment).	1591..2385 1..795	795/795 (100%) 795/795 (100%)	0.0

PFam analysis predicts that the NOV21a protein contains the domains shown in the Table 21E.

Table 21E. Domain Analysis of NOV21a			
Pfam Domain	**NOV21a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
PFK	1028..1039	7/12 (58%) 10/12 (83%)	0.52

Example 22.

The NOV22 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis			
	SEQ ID NO: 91	4170 bp	
NOV22a, CG146119-01 DNA Sequence	ACGCGTTCCCCGGGAACAACACGTTGAGGGCGCCCACCCTGCGTGCCCGGGGCCACCCGGTCCCTGC CCTCGGGCGGCAGGAGAGGTCGAGCTTCCACGGCCCTCGGAGTAGCCCCGTGACCAGACCCGGACTG GCCTTGGAGTTGAAGGGGTTGTTTGCCACCAAATGAACCGAAAAAACTGAACTTTTCAGACTTCGGA GCCTTGGAGTTGAAGGGGTTGTTTGCCACCAAATGAACCGAAAAAACTGAACTTTTCAGACTTCGGA ATGGCAGATATGGGCTTGGAGTTCAGTGAAATTGCAGCGGAGGCGGTGGTGTTCTGAGCTGAG**ATGC** GGCTGCTCCTGCTCGTGCCGCTGCTGCTGGCTCCAGCGCCCGGGTCCTCGGCTCCCAAGGTGAGGCG GCAGAGTGACACCTGGGGACCCTGGAGCCAGTGGAGCCCCTGCAGCCGGACCTGTGGAGGGGGTGTC AGCTTCCGGGAGCGCCCCTGCTACTCCCAGAGGAGAGATGGAGGCTCCAGCTGCGTGGGCCCCGCCC GGAGCCACCGCTCTTGTCGCACGGAGAGCTGCCCCGACGGCGCCCGGGACTTCCGGGCCGAGCAGTG CGCGGAGTTCGACGGAGCGGAGTTCCAGGGGCGGCGGTATCGGTGGCTGCCCTACTACAGCGCCCCA AACAAGTGTGAACTGAACTGCATTCCCAAGGGGGAGAACTTCTACTACAAGCACAGGGAGGCTGTGG TTGATGGGACGCCCTGCGAGCCTGGCAAGAGGGATGTCTGTGTGGATGGCAGCTGCCGGGTTGTCGG CTGTGATCACGAGCTGGACTCGTCCAAGCAGGAGGACAAGTGTCTGCGGTGTGGGGTGACGGCACG ACCTGCTACCCCGTCGCAGGCACCTTTGACGCTAATGACCTCAGCCGAGGCTACAACCAGATCCTCA TAGTTCCCATGGGTGCCACCAGCATCCTCATCGACGAGGCTGCTGCCAGCAGGAACTTCCTGGCTGT GAAGAATGTTCGTGGGGAATACTACCTCAATGGGCACTGGACCATCGAGGCGGCCCGGGCCCTGCCA GCAGCCAGCACCATCCTGCATTACGAGCGGGGTGCTGAGGGGGACCTGGCCCCTGAGCGACTCCATG CCCGGGGCCCCACCTCGGAGCCCCTGGTCATCGAGCTCATCAGCCAGGAGCCCAACCCCGGTGTGCA CTATGAGTACCACCTGCCCCTGCGCCGCCCCAGCCCCGGCTTCAGCTGGAGCCACGGCTCATGGAGT GACTGCAGCGCGGAGTGTGGCGGAGGTCACCAGTCCCGCCTGGTGTTCTGCACCATCGACCATGAGG CCTACCCCGACCACATGTGCCAGCGCCAGCCACGGCCAGCTGACCGGCGTTCCTGCAATCTTCACCC TTGCCCGGAGACCAAGCGCTGGAAGGCAGGGCCATGGGCACCCTGCTCAGCCTCCTGTGGAGGAGGC TCCCAGTCCCGCTCCGTGTACTGCATCTCGTCTGACGGGGCCGGCATCCAGGAGGCCGTGGAGGAGG CTGAGTGTGCCGGGCTGCCTGGGAAGCCCCCTGCCATTCAGGCCTGTAACCTGCAGCGCTGTGCAGC CTGGAGCCCGGAGCCCTGGGGAGAGTGTTCTGTCAGTTGTGGCGTTGGCGTCCGGAAGCGGAGCGTT ACTTGCCGGGGTGAAAGGGGTTCTTTGCTCCATACCGCAGCGTGCTCCTTGGAAGACCGGCCACCTC TGACTGAGCCCTGTGTGCATGAGGACTGCCCCCTCCTCAGTGACCAGGCCTGGCATGTTGGCACCTG GGGTCTATGCTCCAAGAGCTGCAGCTCGGGCACTCGGAGGCGACAGGTCATCTGTGCCATTGGCCG CCCAGCCACTGCGGGAGCCTGCAGCACTCCAAGCCTGTGGATGTGGAGCCTTGTAACACGCAGCCCT GTCATCTCCCCCAGGAGGTCCCCAGCATGCAGGATGTGCACACCCCTGCCAGCAACCCCTGGATGCC GTTGGGCCCTCAGGAGTCCCCTGCCTCAGACTCCAGAGGCCAGTGGTGGGCAGCCCAGGAACACCCC TCAGCCAGGGGTGACCACAGGGGAGAACGAGGTGACCCCAGGGGCGACCAAGGCACCCACCTGTCAG CCCTGGGCCCCGCTCCCTCTCTGCAGCAGCCCCCATACCAGCAACCCCTGCGGTCGGGCTCAGGGCC CCACGACTGCAGACACAGTCCTCACGGGTGCTGCCCCGATGGCCACACGGCATCTCTCGGGCCTCAG TGGCAAGGCTGCCCTGGGGCCCCCTGTCAGCAGAGCAGGTACGGGTGCTGCCCTGACAGGGTATCTG TCGCTGAGGGGCCCCATCACGCTGGCTGCACAAAGTCGTATGGTGGTGACAGCACCGGGGGCATGCC CAGGTCAAGGGCAGTGGCTTCTACAGTAAGTGTCTGGAACACCCACCAGCCCCAGGCCCAGCAGAAT GAGCCCAGTGAGTGCCGGGGCTCCCAGTTTGGCTGTTGCTATGACAACGTGGCCACTGCAGCCGGTC CTCTTGGGGAAGGCTGTGTGGGCCAGCCCAGCCATGCCTACCCCGTGCGGTGCCTGCTGCCCAGTGC CCATGGCTCTTGCGCAGACTGGGCTGCCCGCTGGTACTTCGTTGCCTCTGTGGGCCAATGTAACCGC TTCTGGTATGGCGGCTGCCATGGCAATGCCAATAACTTTGCCTCGGAGCAAGAGTGCATGAGCAGCT GCCAGGGATCTCTCCATGGGCCCCGTCGTCCCCAGCCTGGGGCTTCTGGAAGGAGCACCCACACGGA TGGTGGCGGCAGCAGTCCTGCAGGCGAGCAGGAACCCAGCCAGCACAGGACAGGGGCCGCGGTGCAG AGAAAGCCCTGGCCTTCTGGTGGTCTCTGGCGGCAAGACCAACAGCCTGGGCCAGGGGAGGCCCCCC ACACCCAGGCCTTTGGAGAATGGCCATGGGGGCAGGAGCTTGGGTCCAGGGCCCCTGGACTGGGTGG AGATGCCGGATCACCAGCGCCACCCTTCCACAGCTCCTCCTACAGGATTAGCTTGGCAGGTGTGGAG CCCTCGTTGGTGCAGGCAGCCCTGGGGCAGTTGGTGCGGCTCTCCTGCTCAGACGACACTGCCCCGG AATCCCAGGCTGCCTGGCAGAAAGATGGCCAGCCCATCTCCTCTGACAGGCACAGGCTGCAGTTCGA	

	CGGATCCCTGATCATCCACCCCCTGCAGGCAGAGGACGCGGGCACCTACAGCTGTGGCAGCACCCGG CCAGGCCGCGACTCCCAGAAGATCCAACTTCGCATCATAGGGGGTGACATGGCCGTGCTGTCTGAGG CTGAGCTGAGCCGCTTCCCTCAGCCCAGGGACCCAGCTCAGGACTTTGGCCAAGCGGGGGCTGCTGG GCCCCTGGGGGCCATCCCCTCTTCACACCCACAGCCTGCAAACAGGCTGCGTTTGGACCAGAACCAG CCCCGGGTGGTGGATGCCAGTCCAGGCCAGCGGATCCGGATGACCTGCCGTGCCGAAGGCTTCCCGC CCCCAGCCATCGAGTGGCAGAGAGATGGGCAGCCTGTCTCTTCTCCCAGACACCAGCTGCAGCCTGA TGGCTCCCTGGTCATTAGCCGAGTGGCTGTAGAAGATGGCGGCTTCTACACCTGTGTCGCTTTCAAT GGGCAGGACCGAGACCAGCGATGGGTCCAGCTCAGAGTTCTGGGGGAGCTGACAATCTCAGGACTGC CCCCTACTGTGACAGTGCCAGAGGGTGATACGGCCAGGCTATTGTGTGTGGTAGCAGGAGAAAGTGT GAACATCAGGTGGTCCAGGAACGGGCTACCTGTGCAGGCTGATGGCCACCGTGTCCACCAGTCCCCA GATGGCACGCTGCTCATTTACAACTTGCGGGCCAGGGATGAGGGCTCCTACACGTGCAGTGCCTACC AGGGGAGCCAGGCAGTCAGCCGCAGCACCGAGGTGAAGGTGGTCTCACCAGCACCCACCGCCCAGCC CAGGGACCCTGGCAGGGACTGCGTCGACCAGCCAGAGCTGGCCAACTGTGATTTGATCCTGCAGGCC CAGCTTTGTGGCAATGAGTATTACTCCAGCTTCTGCTGTGCCAGCTGTTCACGTTTCCAGCCTCACG CTCAGCCCATCTGGCAGTAGGGATGAAGGCTAGTTCCAGCCCCAGTCCAAAATAGTTCATAGGGCTA GGGAGAAAGGAAGATG		
	ORF Start: ATG at 265		ORF Stop: TAG at 4105

	SEQ ID NO: 92	1280 aa	MW at 137933.8kD
NOV22a, CG146119-01 Protein Sequence	MRLLLLVPLLLAPAPGSSAPKVRRQSDTWGPWSQWSPCSRTCGGGVSFRERPCYSQRRDGGSSCVGP ARSHRSCRTESCPDGARDFRAEQCAEFDGAEFQGRRYRWLPYYSAPNKCELNCIPKGENFYYKHREA VVDGTPCEPGKRDVCVDGSCRVVGCDHELDSSKQEDKCLRCGGDGTTCYPVAGTFDANDLSRGYNQI LIVPMGATSILIDEAAASRNFLAVKNVRGEYYLNGHWTIEAARALPAASTILHYERGAEGDLAPERL HARGPTSEPLVIELISQEPNPGVHYEYHLPLRRPSPGFSWSHGSWSDCSAECGGGHQSRLVFCTIDH EAYPDHMCQRQPRPADRRSCNLHPCPETKRWKAGPWAPCSASCGGGSQSRSVYCISSDGAGIQEAVE EAECAGLPGKPPAIQACNLQRCAAWSPEPWGECSVSCGVGVRKRSVTCRGERGSLLHTAACSLEDRP PLTEPCVHEDCPLLSDQAWHVGTWGLCSKSCSSGTRRRQVICAIGPPSHCGSLQHSKPVDVEPCNTQ PCHLPQEVPSMQDVHTPASNPWMPLGPQESPASDSRGQWWAAQEHPSARGDHRGERGDPRGDQGTHL SALGPAPSLQQPPYQQPLRSGSGPHDCRHSPHGCCPDGHTASLGPQWQGCPGAPCQQSRYGCCPDRV SVAEGPHHAGCTKSYGGDSTGGMPRSRAVASTVSVWNTHQPQAQQNEPSECRGSQFGCCYDNVATAA GPLGEGCVGQPSHAYPVRCLLPSAHGSCADWAARWYFVASVGQCNRFWYGGCHGNANNFASEQECMS SCQGSLHGPRRPQPGASGRSTHTDGGGSSPAGEQEPSQHRTGAAVQRKPWPSGGLWRQDQQPGPGEA PHTQAFGEWPWGQELGSRAPGLGGDAGSPAPPFHSSSYRISLAGVEPSLVQAALGQLVRLSCSDDTA PESQAAWQKDGQPISSDRHRLQFDGSLIIHPLQAEDAGTYSCGSTRPGRDSQKIQLRIIGGDMAVLS EAELSRFPQPRDPAQDFGQAGAAGPLGAIPSSHPQPANRLRLDQNQPRVVDASPGQRIRMTCRAEGF PPPAIEWQRDGQPVSSPRHQLQPDGSLVISRVAVEDGGFYTCVAFNGQDRDQRWVQLRVLGELTISG LPPTVTVPEGDTARLLCVVAGESVNIRWSRNGLPVQADGHRVHQSPDGTLLIYNLRARDEGSYTCSA YQGSQAVSRSTEVKVVSPAPTAQPRDPGRDCVDQPELANCDLILQAQLCGNEYYSSFCCASCSRFQP HAQPIWQ		

Further analysis of the NOV22a protein yielded the following properties shown in Table 22B.

Table 22B. Protein Sequence Properties NOV22a	
PSort analysis:	0.4896 probability located in outside; 0.1800 probability located in nucleus; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 19 and 20

A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 22C.

Table 22C. Geneseq Results for NOV22a				
Geneseq	Protein/Organism/Length	NOV22a	Identities/	Expect

Identifier	[Patent #, Date]	Residues/ Match Residues	Similarities for the Matched Region	Value
AAU12196	Human PRO4799 polypeptide sequence - Homo sapiens, 477 aa. [WO200140466-A2, 07-JUN-2001]	899..1280 94..477	375/384 (97%) 377/384 (97%)	0.0
ABB71150	Drosophila melanogaster polypeptide SEQ ID NO 40242 - Drosophila melanogaster, 2858 aa. [WO200171042-A2, 27-SEP-2001]	27..538 18..536	209/529 (39%) 282/529 (52%)	e-110
ABB58064	Drosophila melanogaster polypeptide SEQ ID NO 984 - Drosophila melanogaster, 3060 aa. [WO200171042-A2, 27-SEP-2001]	27..538 18..536	209/529 (39%) 282/529 (52%)	e-110
AAU72890	Human metalloprotease partial protein sequence #2 - Homo sapiens, 1103 aa. [WO200183782-A2, 08-NOV-2001]	29..549 550..1064	201/529 (37%) 272/529 (50%)	e-105
AAB74945	Human ADAM type metal protease MDTS2 protein SEQ ID NO:10 - Homo sapiens, 1103 aa. [JP2001008687-A, 16-JAN-2001]	29..549 550..1064	201/529 (37%) 272/529 (50%)	e-105

In a BLAST search of public sequence datbases, the NOV22a protein was found to have homology to the proteins shown in the BLASTP data in Table 22D.

Table 22D. Public BLASTP Results for NOV22a				
Protein Accession Number	**Protein/Organism/Length**	**NOV22a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
O95428	Hypothetical 133.5 kDa protein - Homo sapiens (Human), 1235 aa.	1..1280 1..1235	1136/1321 (85%) 1146/1321 (85%)	0.0
Q9EPX2	Papilin - Mus musculus (Mouse), 1280 aa.	5..1280 6..1278	952/1283 (74%) 1052/1283 (81%)	0.0

Q99JQ8	Hypothetical 52.5 kDa protein - Mus musculus (Mouse), 484 aa.	803..1280 1..482	340/483 (70%) 382/483 (78%)	0.0
Q9U8G8	Lacunin precursor - Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm), 3198 aa.	29..538 63..582	211/530 (39%) 288/530 (53%)	e-113
Q9VAV4	CG1540 protein - Drosophila melanogaster (Fruit fly), 3060 aa.	27..538 18..536	209/529 (39%) 282/529 (52%)	e-109

PFam analysis predicts that the NOV22a protein contains the domains shown in the Table 22E.

Table 22E. Domain Analysis of NOV22a			
Pfam Domain	**NOV22a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
tsp_1	30..79	23/55 (42%) 39/55 (71%)	3.7e-10
tsp_1	309..360	14/58 (24%) 34/58 (59%)	0.042
tsp_1	366..424	15/65 (23%) 39/65 (60%)	0.0013
tsp_1	425..480	19/61 (31%) 40/61 (66%)	0.0013
tsp_1	488..538	14/54 (26%) 36/54 (67%)	0.0095
Kunitz_BPTI	756..806	24/62 (39%) 41/62 (66%)	2e-27
ig	926..982	17/60 (28%) 41/60 (68%)	2.9e-07
ig	1060..1116	20/60 (33%) 45/60 (75%)	4.1e-12
ig	1149..1206	15/62 (24%) 42/62 (68%)	1.2e-09

Example 23.

The NOV23 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis			
	SEQ ID NO: 93	1088 bp	
NOV23a, CG146202-01 DNA Sequence	CTCCTGGAGGAGGAACAGCTGAGAGGCCTTGGATTCCGACAGACTCGAGGATACAAGAGCTTAGCAG GGTGTCTTGGCCATGGTCCCCTGGTGCTGCAACTCCTCTCCTTCACGCTCTTGGCTGGGCTCCTTGT CCAAGTGTCCAAGGTCCCCAGCTCCATAAGTCAGGAACAATCCAGGCAAGACGCGATCTACCAGAAC CTGACCCAGCTTAAAGCTGCAGTGGGTGAGCTCTCAGAGAAATCCAAGCTGCAGGAGATCTACCAGG AGCTGACCCAGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCTGCAGGAGATCTACCA GGAGCTGACCCGGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCTGCAGGAGATCTAC CAGGAGCTGACCTGGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGATGCAGGAGATCT ACCAGGAGCTGACTCGGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCAGCAGGAGAT CTACCAGGAGCTGACCCGGCTGAAGGCTGCAGTGGAACGCCTGTGCCACCCCTGTCCCTGGGAATGG ACATTCTTCCAAGGAAACTGTTACTTCATGTCTAACTCCCAGCGGAACTGGCACGACTCCATCACCG CCTGCAAAGAAGTGGGGGCCCAGCTCGTCGTAATCAAAAGTGCTGAGGAGCAGAACTTCCTACAGCT GCAGTCTTCCAGAAGTAACCGCTTCACCTGGATGGGACTTTCAGATCTAAATCAGGAAGGCACGTGG CAATGGGTGGACGGCTCACCTCTGTTGCCCAGCTTCAAGCAGTATTGGAACAGAAGAGAGCCCAACA ACGTTGGGGAGGAAGACTGCGCGGAATTTAGTGGCAATGGCTGGAACGACGACAAATGTAATCTTGC CAAATTCTGGATCTGCAAAAAGTCCGCAGCCTCCTGCTCCAGGGATGAAGAACAGTTTCTTTCTCCA GCCCCTGCCACCCCAAACCCCCCTCCTGCG**TAG**CAGAACTTCACCCCCTTTTAAGCTACAGTTCCTT CTCTCCATCCTTCGAC		
	ORF Start: at 1		ORF Stop: TAG at 1036

	SEQ ID NO: 94	345 aa	MW at 39149.0kD
NOV23a, CG146202-01 Protein Sequence	LLEEEQLRGLGFRQTRGYKSLAGCLGHGPLVLQLLSFTLLAGLLVQVSKVPSSISQEQSRQDAIYQN LTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPEKSKLQEIYQELTRLKAAVGELPEKSKLQEIY QELTWLKAAVGELPEKSKMQEIYQELTRLKAAVGELPEKSKQQEIYQELTRLKAAVERLCHPCPWEW TFFQGNCYFMSNSQRNWHDSITACKEVGAQLVVIKSAEEQNFLQLQSSRSNRFTWMGLSDLNQEGTW QWVDGSPLLPSFKQYWNRREPNNVGEEDCAEFSGNGWNDDKCNLAKFWICKKSAASCSRDEEQFLSP APATPNPPPA		

Further analysis of the NOV23a protein yielded the following properties shown in Table 23B.

Table 23B. Protein Sequence Properties NOV23a	
PSort analysis:	0.7900 probability located in plasma membrane; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in Golgi body; 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 50 and 51

A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 23C.

Table 23C. Geneseq Results for NOV23a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAU88025	Mouse OtS1-B7 ORF protein - Mus sp, 404 aa. [WO200214366-A2, 21-FEB-2002]	1..345 14..404	344/391 (87%) 344/391 (87%)	0.0
AAG79086	Human DC-SIGN, a dendritic cell-speciific C-type lectin - Homo sapiens, 404 aa. [WO200164752-A2, 07-SEP-2001]	1..345 14..404	344/391 (87%) 344/391 (87%)	0.0
AAB28614	Human C-type lectin receptor - Homo sapiens, 404 aa. [WO200063251-A1, 26-OCT-2000]	1..345 14..404	344/391 (87%) 344/391 (87%)	0.0
AAB19714	Dendritic cell specific C-type lectin DC-SIGN - Homo sapiens, 404 aa. [EP1046651-A1, 25-OCT-2000]	1..345 14..404	344/391 (87%) 344/391 (87%)	0.0
AAR32188	Sequence of a non-CD4 glycoprotein gp120 receptor protein - Homo sapiens, 404 aa. [WO9301820-A, 04-FEB-1993]	1..345 14..404	338/391 (86%) 340/391 (86%)	0.0

In a BLAST search of public sequence datbases, the NOV23a protein was found to have homology to the proteins shown in the BLASTP data in Table 23D.

Table 23D. Public BLASTP Results for NOV23a

Protein Accession Number	Protein/Organism/Length	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NNX6	Membrane-associated lectin type-C (Probable mannose-binding C-type lectin DC-SIGN) (MDC-SIGN1A type I isoform) - Homo sapiens (Human), 404 aa.	1..345 14..404	344/391 (87%) 344/391 (87%)	0.0
Q96QQ1	MDC-SIGN1B type I isoform - Homo sapiens (Human), 404 aa.	3..345 16..404	342/389 (87%) 342/389 (87%)	0.0
Q96QQ8	MDC-SIGN1A type II isoform - Homo sapiens (Human), 398 aa.	1..345 14..398	338/391 (86%) 338/391 (86%)	0.0

Q95LC6	Dendritic cell-specific ICAM-3 grabbing nonintegrin - Macaca nemestrina (Pig-tailed macaque), 381 aa.	1..345 14..381	319/368 (86%) 332/368 (89%)	0.0
Q95LA8	Dendritic cell-specific ICAM-3 grabbing nonintegrin - Macaca mulatta (Rhesus macaque), 381 aa.	1..345 14..381	317/368 (86%) 331/368 (89%)	0.0

PFam analysis predicts that the NOV23a protein contains the domains shown in the Table 23E.

Table 23E. Domain Analysis of NOV23a

Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
integrin_B	65..83	13/21 (62%) 19/21 (90%)	0.25
lectin_c	214..320	45/127 (35%) 87/127 (69%)	3.6e-34

Example 24.

The NOV24 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 24A.

Table 24A. NOV24 Sequence Analysis

	SEQ ID NO: 95	1191 bp	
NOV24a, CG146250-01 DNA Sequence	GAGTACGTATCGTCCACTCTGAGCCTTAGAGGTGGGGGTTCATCAGGAGCACTTCGAGGAGGAGGAG GAGGAGGCCGGGGTGGAGGGGTGGGCTCTGGCGGCCTCTGTCGAGCCCTCCGCTCCTATGCGCTCTG CACTCGGCGCACCGCCCGCACCTGCCGCGGGGACCTCGCCTTCCATTCGGCGGTACATGGCATCGAA GACCTGATGATCCAGCACAACTGCTCCCGCCAGGGCCCTACAGCCCCTCCCCCGCCCCGGGGCCCCG CCCTTCCAGGCGCGGCTCCGGCCTCCCTGCCCCGGACCCTTGTGACTATGAAGGCCGGTTTTCCCG GCTGCATGGTCGTCCCCCGGGGTTCTTGCATTGCGCTTCCTTCGGGGACCCCCATGTGCGCAGCTTC CACCATCACTTTCACACATGCCGTGTCCAAGGAGCTTGGCCTCTACTGGATAATGACTTCCTCTTTG TCCAAGCCACCAGCTCCCCCATGGCGTTGGGGGCCAACGCTACCGCCACCCGGAAGGTCACCATCAT ATTTAAGAACATGCAGGAATGCATTGATCAGAAGGTGTATCAGGCTGAGGTGGATAATCTTCCTGTA GCCTTTGAAGATGGTTCTATCAATGGAGGTGACCGACCTGGGGGATCCAGTTTGTCGATTCAAACTG CTAACCCTGGGAACCATGTGGAGATCCAAGCTGCCTACATTGGCACAACTATAATCATTCGGCAGAC AGCTGGGCAGCTCTCCTTCTCCATCAAGGTAGCAGAGGATGTGGCCATGGCCTTCTCAGCTGAACAG GACCTGCAGCTCTGTGTTGGGGGGTGCCCTCCAAGTCAGCGACTCTCTCGATCAGAGCGCAATCGTC GGGGAGCTATAACCATTGATACTGCCAGACGGCTGTGCAAGGAAGGGCTTCCAGTGGAAGATGCTTA CTTCCATTCCTGTGTCTTTGATGTTTTAATTTCTGGTGATCCCAACTTTACCGTGGCAGCTCAGGCA GCACTGGAGGATGCCCGAGCCTTCCTGCCAGACTTAGAGAAGCTGCATCTCTTCCCCTCAGATGCTG GGGTTCCTCTTTCCTCAGCAACCCTCTTAGCTCCACTCCTTTCTGGGCTCTTTGTTCTGTGGCTTTG CATTCAGTAAGGGGACCATCAGTCCCATTACTAGTTTGGAAATGATTTGGAG		
	ORF Start: ATG at 208		ORF Stop: TAA at 1147

	SEQ ID NO: 96	313 aa	MW at 33664.8kD

NOV24a, CG146250-01 Protein Sequence	MIQHNCSRQGPTAPPPPRGPALPGAGSGLPAPDPCDYEGRFSRLHGRPPGFLHCASFGDPHVRSFHH HFHTCRVQGAWPLLDNDFLFVQATSSPMALGANATATRKVTIIFKNMQECIDQKVYQAEVDNLPVAF EDGSINGGDRPGGSSLSIQTANPGNHVEIQAAYIGTTIIIRQTAGQLSFSIKVAEDVAMAFSAEQDL QLCVGGCPPSQRLSRSERNRRGAITIDTARRLCKEGLPVEDAYFHSCVFDVLISGDPNFTVAAQAAL EDARAFLPDLEKLHLFPSDAGVPLSSATLLAPLLSGLFVLWLCIQ

	SEQ ID NO: 97	974 bp	
NOV24b, CG146250-02 DNA Sequence	AAGACCTGATGATCCAGCACAACTGCTCCCGCCAGGGCCCTACAGCCCCTCCCCCGCCCCGGGCCC CGCCCTTCCAGGCGCGGCTCCGGCCTCCCTGCCCCGGACCCTTGTGACTATGAAGGCCGGTTTTCC CGGCTGCATGGTCGTCCCCCGGGGTTCTTGCATTGCGCTTCCTTCGGGGACCCCCATGTGCGCAGCT TCCACCATCACTTTCACACATGCCGTGTCCAAGGAGCTCGGCCTCTACTGGATAATGACTTCCTCTT TGTCCAAGCCACCAGCTCCCCCATGGCGTTGGGGGCCAACGCTACCGCCACCCGGAAGCTCACCATC ATATTTAAGAACATGCAGGAATGCATTGATCAGAAGGTGTATCAGGCTGAGGTGGATAATCTTCCTG TAGCCTTTGAAGATGGTTCTATCAATGGAGGTGACCGACCTGGGGGATCCAGTTTGTCGATTCAAAC TGCTAACCCTGGGAACCATGTGGAGATCCAAGCTGCCTACATTGGCACAACTATAATCATTCGGCAG ACAGCTGGGCAGCTCTCCTTCTCCATCAAGGTAGCAGAGGATGTGGCCATGGCCTTCTCAGCTGAAC AGGACCTGCAGCTCTGTGTTGGGGGGTGCCCTCCAAGTCAGCGACTCTCTCGATCAGAGCGCAATCG TCGGGGAGCTATAACCATTGATACTGCCAGACGGCTGTGCAAGGAAGGGCTTCCAGTGGAAGATGCT TACTTCCATTCCTGTGTCTTTGATGTTTTAATTTCTGGTGATCCCAACTTTACCGTGGCAGCTCAGG CAGCACTGGAGGATGCCCGAGCCTTCCTGCCAGACTTAGAGAAGCTGCATCTCTTCCCCTCAGATGC TGGGGTTCCTCTTTCCTCAGCAACCCTCTTAGCTCCACTCCTTTCTGGGCTCTTTGTTCTGTGGCTT TGCATTCAGTAAGGGGAACCATCAGTACAGGGCGAT		
	ORF Start: ATG at 9		ORF Stop: TAA at 948

	SEQ ID NO: 98	313 aa	MW at 33648.8kD
NOV24b, CG146250-02 Protein Sequence	MIQHNCSRQGPTAPPPPRGPALPGAGSGLPAPDPCDYEGRFSRLHGRPPGFLHCASFGDPHVRSFHH HFHTCRVQGARPLLDNDFLFVQATSSPMALGANATATRKLTIIFKNMQECIDQKVYQAEVDNLPVAF EDGSINGGDRPGGSSLSIQTANPGNHVEIQAAYIGTTIIIRQTAGQLSFSIKVAEDVAMAFSAEQDL QLCVGGCPPSQRLSRSERNRRGAITIDTARRLCKEGLPVEDAYFHSCVFDVLISGDPNFTVAAQAAL EDARAFLPDLEKLHLFPSDAGVPLSSATLLAPLLSGLFVLWLCIQ		

	SEQ ID NO: 99	1338 bp	
NOV24c, CG146250-03 DNA Sequence	CCGGCGCCTGGGAAACCTGGCTGGATAGGTATGGGGGAGCCAGGCCAGTCCCCTAGTCCCAGGTCCT CCCATGGCAGTCCCCCAACTCTAAGCACTCTCACTCTCCTGCTGCTCCTCTGTGGATTAGCTCATTC TCAATGCAAGATCCTCCGCTGCAATGCTGAGTACGTATCGTCCACTCTGAGCCTTAGAGGTGGGGT TCATCAGGAGCACTTCGAGGAGGAGGAGGAGGAGGCCGGGGTGGAGGGGTGGGCTCTGGCGGCCTCT GTCGAGCCCTCCGCTCCTATGCGCTCTGCACTCGGCGCACCGCCCGCACCTGCCGCGGGGACCTCGC CTTCCATTCGGCGGTACATGGCATCGAAGACCTGATGATCCAGCACAACTGCTCCCGCCAGGGCCCT ACAGCCCCTCCCCCGCCCCGGGGCCCCGCCCTTCCAGGCGCGGCTCCGGCCTCCCTGCCCCGGACC CTTGTGACTATGAAGGCCGGTTTTCCCGGCTGCATGGTCGTCCCCCGGGGTTCTTGCATTGCGCTTC CTTCGGGGACCCCCATGTGCGCAGCTTCCACCATCACTTTCACACATGCCGTGTCCAAGGAGCTTGG CCTCTACTGGATAATGACTTCCTCTTTGTCCAAGCCACCAGCTCCCCCATGGCGTTGGGGGCCAACG CTACCGCCACCCGGAAGGTCACCATCATATTTAAGAACATGCAGGAATGCATTGATCAGAAGGTGTA TCAGGCTGAGGTGGATAATCTTCCTGTAGCCTTTGAAGATGGTTCTATCAATGGAGGTGACCGACCT GGGGGATCCAGTTTGTCGATTCAAACTGCTAACCCTGGGAACCATGTGGAGATCCAAGCTGCCTACA TTGGCACAACTATAATCATTCGGCAGACAGCTGGGCAGCTCTCCTTCTCCATCAAGGTAGCAGAGGA TGTGGCCATGGCCTTCTCAGCTGAACAGGACCTGCAGCTCTGTGTTGGGGGGTGCCCTCCAAGTCAG CGACTCTCTCGATCAGAGCGCAATCGTCGGGGAGCTATAACCATTGATACTGCCAGACGGCTGTGCA AGGAAGGGCTTCCAGTGGAAGATGCTTACTTCCATTCCTGTGTCTTTGATGTTTTAATTTCTGGTGA TCCCAACTTTACCGTGGCAGCTCAGGCAGCACTGGAGGATGCCCGAGCCTTCCTGCCAGACTTAGAG AAGCTGCATCTCTTCCCCTCAGATGCTGGGGTTCCTCTTTCCTCAGCAACCCTCTTAGCTCCACTCC TTTCTGGGCTCTTTGTTCTGTGGCTTTGCATTCAGTAAGGGGACCATCAGTCCCATTACTAGTTT		
	ORF Start: ATG at 31		ORF Stop: TAA at 1309

	SEQ ID NO: 100	426 aa	MW at 45041.5kD
NOV24c, CG146250-03 Protein Sequence	MGEPGQSPSPRSSHGSPPTLSTLTLLLLLCGLAHSQCKILRCNAEYVSSTLSLRGGGSSGALRGGGG GGRGGGVGSGGLCRALRSYALCTRRTARTCRGDLAFHSAVHGIEDLMIQHNCSRQGPTAPPPPRGPA LPGAGSGLPAPDPCDYEGRFSRLHGRPPGFLHCASFGDPHVRSFHHHFHTCRVQGAWPLLDNDFLFV QATSSPMALGANATATRKVTIIFKNMQECIDQKVYQAEVDNLPVAFEDGSINGGDRPGGSSLSIQTA NPGNHVEIQAAYIGTTIIIRQTAGQLSFSIKVAEDVAMAFSAEQDLQLCVGGCPPSQRLSRSERNRR GAITIDTARRLCKEGLPVEDAYFHSCVFDVLISGDPNFTVAAOAALEDARAFLPDLEKLHLFPSDAG		

	VPLSSATLLAPLLSGLFVLWLCIQ

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 24B.

Table 24B. Comparison of NOV24a against NOV24b and NOV24c.		
Protein Sequence	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV24b	1..313 1..313	271/313 (86%) 272/313 (86%)
NOV24c	1..313 114..426	273/313 (87%) 273/313 (87%)

Further analysis of the NOV24a protein yielded the following properties shown in Table 24C.

Table 24C. Protein Sequence Properties NOV24a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.3740 probability located in microbody (peroxisome); 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 24D.

Table 24D. Geneseq Results for NOV24a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU70169	Rat secreted factor protein encoded by DNA clone P0248_B04 - Rattus norvegicus, 422 aa. [WO200174901-A2, 11-OCT-2001]	1..313 110..422	273/313 (87%) 284/313 (90%)	e-161
AAM93823	Human polypeptide, SEQ ID NO: 3881 - Homo sapiens, 450 aa. [EP1130094-A2, 05-SEP-2001]	1..308 110..446	158/340 (46%) 206/340 (60%)	1e-76

ABG65106	Human albumin fusion protein #1781 - Homo sapiens, 434 aa. [WO200177137-A1, 18-OCT-2001]	1..308 94..430	157/340 (46%) 205/340 (60%)	4e-76
AAE07112	Human gene 6 encoded secreted protein fragment, SEQ ID NO:129 - Homo sapiens, 471 aa. [WO200154708-A1, 02-AUG-2001]	1..308 131..467	157/340 (46%) 205/340 (60%)	4e-76
AAE07056	Human gene 6 encoded secreted protein HARMJ38, SEQ ID NO:73 - Homo sapiens, 434 aa. [WO200154708-A1, 02-AUG-2001]	1..308 94..430	157/340 (46%) 205/340 (60%)	4e-76

In a BLAST search of public sequence datbases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24E.

Table 24E. Public BLASTP Results for NOV24a

Protein Accession Number	Protein/Organism/Length	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
BAC03944	CDNA FLJ35363 fis, clone SKMUS2000679 - Homo sapiens (Human), 313 aa.	1..313 1..313	311/313 (99%) 313/313 (99%)	0.0
BAC05248	CDNA FLJ40846 fis, clone TRACH2014544 - Homo sapiens (Human), 422 aa.	1..313 110..422	273/313 (87%) 284/313 (90%)	e-161
Q8WVJ5	Similar to RIKEN cDNA 2310035L15 gene - Homo sapiens (Human), 200 aa.	114..313 1..200	200/200 (100%) 200/200 (100%)	e-111
AAH22603	Hypothetical protein - Mus musculus (Mouse), 201 aa.	114..311 1..198	177/198 (89%) 182/198 (91%)	1e-97
Q9D741	2310035L15Rik protein - Mus musculus (Mouse), 201 aa.	114..311 1..198	177/198 (89%) 182/198 (91%)	2e-97

PFam analysis predicts that the NOV24a protein contains the domains shown in the Table 24F.

Table 24F. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 25.

The NOV25 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 25A.

Table 25A. NOV25 Sequence Analysis			
	SEQ ID NO: 101	381 bp	
NOV25a, CG146625-01 DNA Sequence	ATTCCTGGTGGTGAAAGGATGTGGCCCCAGGACCCATCCCGGAAGGAGGTGCTGAGGTTTGCAGTCA GCTGCCGTATCCTGACTCTGATGCTGCAGGTTCTCACCAGGTTTTGGGCTCCTCCACTCCTATTAT GTACTGGTTTCCAGCTCACTTGCTTCAGGATCAAGAGCCGCTGTTGAGATCCTTAAAGACTGTGCCT TGGAAGCCTCTTGCAGAGGACTCCCCACCAGGACAAAAGGTCCCCAGAAATCCTATCATGGGACTTT TGTATCACTGGAAAACCTGTTCTCCAGTCACACGATACATTCTAGGCTACTTCCTGACTTACTGGCT CCTGGGACTACTCCTACATTGCAACTTCCTGCCTTGGACATGACCT		
	ORF Start: ATG at 19		ORF Stop: TGA at 376

	SEQ ID NO: 102	119 aa	MW at 13984.5kD
NOV25a, CG146625-01 Protein Sequence	MWPQDPSRKEVLRFAVSCRILTLMLQVLTRFLGSSTPIMYWFPAHLLQDQEPLLRSLKTVPWKPLAE DSPPGQKVPRNPIMGLLYHWKTCSPVTRYILGYFLTYWLLGLLLHCNFLPWT		

	SEQ ID NO: 103	906 bp	
NOV25b, CG146625-02 DNA Sequence	GGAGCTCAATCCTGGTAGCAACACCCCTGAATTCCTGGTGGTGAAAGGATGTGGCCCCAGGACCCAT CCCGGAAGGAGGTGCTGAGGTTTGCAGTCAGCTGCCGTATCCTGACTCTGATGCTGCAGGCCCTCTT CAATGCCATCATCCCAGATCACCATGCAGAAGCCTTCTCTCCTCCTCGCCTGGCCCCCTCAGGCTTT GTGGACCAACTCGTGGAAGGCTCAGCCCGCCCCATTCCTGAGCCTTTGGTACAGTTAGCTGTAGACA AGGGCTACCGGATTGCAGAGGGAAATGAACCGCCTTGGTGCTTCTGGGATGTTCCACTAATATACAG CTATATCCAGGATGTCTGCTGGAATGTTGGCTTTTTGAAATACTATGAGCTCAAGCAGGTGCCCAAT TTTCTACTGGCTGCACCAGTGGCTATACTGGTTGCCTGGGCAACTTGGACATACGTGACCACTCACC CTTGGCTCTGCCTTACACTTGGGCTGCAAAGGAGCAAGAACAATAAGACCCTAGAGAAGCCCGATCT TGGATTCCTCAGTCCTCAGGTGTTTGTGTACGTGGTCCACGCTGCAGTGCTGCTGCTGTTTGGAGGT CTGTGCATGCATGTTCAGGTTCTCACCAGGTTTTTGGGCTCCTCCACTCCTATTATGTACTGGTTTC CAGCTCACTTGCTTCAGGATCAAGAGCCGCTGTTGAGATCCTTAAAGACTGTGCCTTGGAAGCCTCT TGCAGAGGACTCCCCACCAGGACAAAAGGTCCCCAGAAATCCTATCATGGGACTTTTGTATCACTGG AAAACCTGTTCTCCAGTCACACGATACATTCTAGGCTACTTCCTGACTTACTGGCTCCTGGGACTAC TCCTACATTGCAACTTCCTGCCTTGGACATGACCT		
	ORF Start: ATG at 49		ORF Stop: TGA at 901

	SEQ ID NO: 104	284 aa	MW at 32499.9kD
NOV25b, CG146625-02 Protein Sequence	MWPQDPSRKEVLRFAVSCRILTLMLQALFNAIIPDHHAEAFSPPRLAPSGFVDQLVEGSARPIPEPL VQLAVDKGYRIAEGNEPPWCFWDVPLIYSYIQDVCWNVGFLKYYELKQVPNFLLAAPVAILVAWATW TYVTTHPWLCLTLGLQRSKNNKTLEKPDLGFLSPQVFVYVVHAAVLLLFGGLCMHVQVLTRFLGSST PIMYWFPAHLLQDQEPLLRSLKTVPWKPLAEDSPPGQKVPRNPIMGLLYHWKTCSPVTRYILGYFLT YWLLGLLLHCNFLPWT		

	SEQ ID NO: 105	2114 bp	
NOV25c, CG146625-03 DNA Sequence	CTCGTCTGCTTCCGGCCCTGTGGCCTGGTGGGGCTCTGCAGGCTCCCTCGGGAGTGGTCCTTGGGCC GTGGCCCCTCTGGGAGGCCTGAGGGAGCTCAATCCTGGTAGCAACACCCCTGAATTCCTGGTGGTGA AAGGATGTGGCCCCAGGACCCATCCCGGAAGGAGGTGCTGAGGTTTGCAGTCAGCTGCCGTATCCTG ACTCTGATGCTGCAGGCCCTCTTCAATGCCATCACCCCAGATCACCATGCAGAAGCCTTCTCTCCTC CTCGCCTGGCCCCCTCAGGCTTTGTGGACCAACTCGTGGAAGGTCTTCTGGGCGGCCTGTCTCACTG GGATGCTGAACACTTCTTGTTCATTGCTGAGCATGGCTACCTGTATGAGCACAACTTTGCCTTCTTT CCTGGTTTCCCCTTGGCCCTGCTGGTGGGGACTGAACTGTTGAGACCCTTACGGGGTTACTGAGTC TACGCAGTTGCCTGCTGATTTCGGTAGCATCACTCAATTTCTTGTTCTTCATGTTGGCTGCAGTTGC ACTTCATGACCTGGGTTGTCTGGTTTTGCACTGTCCCCACCAGTCCTTTTATGCAGCTCTGCTTTTC TGTCTCAGCCCTGCCAATGTCTTCCTGGCAGCTGGTTACTCAGAAGCTTTGTTTGCCCTCCTGACAT TCAGTGCCATGGGGCAGCTGGAGAGGGGCCGAGTCTGGACTAGTGTACTCCTCTTTGCCTTTGCCAC TGGGGTACGCTCCAACGGGCTGGTCAGTGTTGGCTTCCTCATGCATTCTCAATGCCAAGGCTTTTTC TCTTCTCTAACGATGCTGAATCCTCTGAGACAGCTCTTTAAGCTGATGGCCTCTCTGTTTCTGTCGG TGTTCACACTTGGCCTTCCCTTTGCCCTCTTTCAGTATTATGCCTACACCCAATTCTGTCTGCCAGG CTCAGCCCGCCCCATTCCTGAGCCTTTGGTACAGTTAGCTGTAGACAAGGGCTACCGGATTGCAGAG GGAAATGAACCGCCTTGGTGCTTCTGGGATGTTCCACTAATATACAGCTATATCCAGGATGTCTACT GGAATGTTGGCTTTTTGAAATACTATGAGCTCAAGCAGGTGCCCAATTTTCTACTGGCTGCACCAGT GGCTATACTGGTTGCCTGGGCAACTTGGACATACGTGACCACTCACCCTTGGCTCTGCCTTACACTT GGGCTGCAAAGGAGCAAGAACAATAAGACCCTAGAGAAGCCCGATCTTGGATTCCTCAGTCCTCAGG TGTTTGTGTACGTGGTCCACGCTGCAGTGCTGCTGCTGTTTGGAGGTCTGTGCATGCATGTTCAGGT TCTCACCAGGTTTTTGGGCTCCTCCACTCCTATTATGTACTGGTTTCCAGCTCACTTGCTTCAGGAT CAAGAGCCGCTGTTGAGATCCTTAAAGACTGTGCCTTGGAAGCCTCTTGCAGAGGACTCCCCACCAG GACAAAAGGTCCCCAGAAATCCTATCATGGGACTTTTGTATCACTGGAAAACCTGTTCTCCAGTCAC ACGATACATTCTAGGCTACTTCCTGACTTACTGGCTCCTGGGACTACTCCTACATTGCAACTTCCTG CCTTGGACATGACCTGGACTCTCCAGGGACAGGTTGGAAGCCAACTTAACCCAGGGGTCTGAAAGTA AAAATACACATTGGAACTGCCTCTGCTGCCCTGGGATCATTACTGTGTCCATTATAATCTTTCTCTT TCTCTTTGAAAGCTGGTCAGGAATGGGAGAAGTGTCAGACACTAGAGAGCCCCTTCTGGTCCTGGCT AGGGCAAATTTTAGACAACTATTTTCTCTGTAAGTGAAGATTGTCGTATTCCAAGTCTAAAATACAC CTGGATCTGTCTAGTCAATCAACATAGCAGAGACAGTCTTAAACCTACCATTGACCTGTGTGTAAAT TTAAATGTCAATTTATTGAAGTGTAAATTTCATCAAAGGCATTAGCTGACAGGCTGGTAACAGTCCA CACAAGATGGTATAGGCCTGAACAGTGTAGTGGCAGTAATAAAGTGGGACCATTTTTTCCAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 139		ORF Stop: TGA at 1618

	SEQ ID NO: 106	493 aa	MW at 55699.9kD
NOV25c, CG146625-03 Protein Sequence	MWPQDPSRKEVLRFAVSCRILTLMLQALFNAITPDHHAEAFSPPRLAPSGFVDQLVEGLLGGLSHWD AEHFLFIAEHGYLYEHNFAFFPGFPLALLVGTELLRPLRGLLSLRSCLLISVASLNFLFFMLAAVAL HDLGCLVLHCPHQSFYAALLFCLSPANVFLAAGYSEALFALLTFSAMGQLERGRVWTSVLLFAFATG VRSNGLVSVGFLMHSQCQGFFSSLTMLNPLRQLFKLMASLFLSVFTLGLPFALFQYYAYTQFCLPGS ARPIPEPLVQLAVDKGYRIAEGNEPPWCFWDVPLIYSYIQDVYWNVGFLKYYELKQVPNFLLAAPVA ILVAWATWTYVTTHPWLCLTLGLQRSKNNKTLEKPDLGFLSPQVFVYVVHAAVLLLFGGLCMHVQVL TRFLGSSTPIMYWFPAHLLQDQEPLLRSLKTVPWKPLAEDSPPGQKVPRNPIMGLLYHWKTCSPVTR YILGYFLTYWLLGLLLHCNFLPWT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 25B.

Table 25B. Comparison of NOV25a against NOV25b and NOV25c.

Protein Sequence	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV25b	21..119 186..284	81/99 (81%) 83/99 (83%)
NOV25c	21..119 395..493	81/99 (81%) 83/99 (83%)

Further analysis of the NOV25a protein yielded the following properties shown in Table 25C.

Table 25C. Protein Sequence Properties NOV25a	
PSort analysis:	0.8025 probability located in lysosome (lumen); 0.7480 probability located in microbody (peroxisome); 0.4715 probability located in mitochondrial matrix space; 0.1742 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 34 and 35

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 25D.

Table 25D. Geneseq Results for NOV25a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV25a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABG65069	Human albumin fusion protein #1744 - Homo sapiens, 280 aa. [WO200177137-A1, 18-OCT-2001]	21..119 182..280	95/99 (95%) 97/99 (97%)	8e-55
ABB89811	Human polypeptide SEQ ID NO 2187 - Homo sapiens, 173 aa. [WO200190304-A2, 29-NOV-2001]	21..119 75..173	95/99 (95%) 97/99 (97%)	8e-55
ABB97380	Novel human protein SEQ ID NO: 648 - Homo sapiens, 493 aa. [WO200222660-A2, 21-MAR-2002]	21..119 395..493	95/99 (95%) 97/99 (97%)	8e-55
AAE07114	Human gene 9 encoded secreted protein fragment, SEQ ID NO:131 - Homo sapiens, 311 aa. [WO200154708-A1, 02-AUG-2001]	21..119 213..311	95/99 (95%) 97/99 (97%)	8e-55
AAE07059	Human gene 9 encoded secreted protein HTEGF16, SEQ ID NO:76 - Homo sapiens, 280 aa. [WO200154708-A1, 02-AUG-2001]	21..119 182..280	95/99 (95%) 97/99 (97%)	8e-55

In a BLAST search of public sequence datbases, the NOV25a protein was found to have homology to the proteins shown in the BLASTP data in Table 25E.

Table 25E. Public BLASTP Results for NOV25a				
Protein Accession Number	Protein/Organism/Length	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NUD9	DJ50O24.5.1 (Novel protein (Translation of cDNA KAT07271 (Em:AK000484))) (Hypothetical 55.7 kDa protein) - Homo sapiens (Human), 493 aa.	21..119 395..493	95/99 (95%) 97/99 (97%)	2e-54
Q9NX26	CDNA FLJ20477 fis, clone KAT07271 - Homo sapiens (Human), 493 aa.	21..119 395..493	95/99 (95%) 97/99 (97%)	2e-54
Q9U3X2	VEGETABLE precursor - Drosophila melanogaster (Fruit fly), 449 aa.	9..119 361..449	38/123 (30%) 50/123 (39%)	0.057
Q9V7W1	CG6657 protein - Drosophila melanogaster (Fruit fly), 449 aa.	9..119 361..449	38/123 (30%) 50/123 (39%)	0.057
Q95TV6	GM14315p - Drosophila melanogaster (Fruit fly), 273 aa.	9..119 185..273	38/123 (30%) 50/123 (39%)	0.057

PFam analysis predicts that the NOV25a protein contains the domains shown in the Table 25F.

Table 25F. Domain Analysis of NOV25a			
Pfam Domain	NOV25a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 26.

The NOV26 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 26A.

Table 26A. NOV26 Sequence Analysis			
	SEQ ID NO: 107	1139 bp	
NOV26a, CG147284-01 DNA Sequence	GCACTCACTACGCACAGACTCGACGGTGCCATCAGCATGAGAACTTACCGCTACTTCTTGCTGCTCT TTTGGGTGGGCCAGCCCTACCCAACTCTCTCAACTCCACTATCAAAGAGGACTAGTGGTTTCCCAGC AAAGAAAAGGGCCCTGGAGCTCTCTGGAAACAGCAAAAATGAGCTGAACCGTTCAAAAAGGAGCTGG ATGTGGAATCAGTTCTTTCTCCTGGAGGAATACACAGGATCCGATTATCAGTATGTGGGCAAGTTAC ATTCAAACTTTACCATTCAAGACAACAAAGACAACACGGCGGGAATCTTAACTCGGAAAAATGGCTA TAATAGACACGAGATGAGCACCTATCTCTTGCCTGTGGTCATTTCAGACAACGACTACCCAGTTCAA AGCAGCACTGGGACAGTGACTGTCCGGGTCTGTGCATGTGACCACCACGGGAACATGCAATCCTGCC ACGCGGAGGCGCTCATCCACCCCACGGGACTGAGCACGGGGGCTCTGGTTGCCATCCTTCTGTGCAT CGTGATCCTACTAGTGACAGTGGTGCTGTTTGCAGCTCTGAGGCGGCAGCGAAAAAAAGAGCCTTTG ATCATTTCCAAAGAGGACATCAGAGATAACATTGTCAGTTACAACGACGAAGGTGGTGGAGAGGAGG ACACCCAGGCTTTTGATATCGGCACCCTGAGGAATCCTGAAGCCATAGAGGACAACAAATTACGAAG GGACATTGTGCCCGAAGCCCTTTTCCTACCCCGACGGACTCCAACAGCTCGCGACAACACCGATGTC AGAGATTTCATTAACCAAAGGTTAAAGGAAAATGACACGGACCCCACTGCCCCGCCATACGACTCCT TGGCCACTTACGCCTATGAAGGCACTGGCTCCGTGGCGGATTCCCTGAGCTCGCTGGAGTCAGTGAC CACGGATGCAGATCAAGACTATGATTACCTTAGTGACTGGGGACCTCGATTCAAAAAGCTTGCAGAT ATGTATGGAGGAGTGGACAGTGACAAAGACTCCTAATCTGTTGCCTTTTTCATTTTCCAATACGACA CTGAAATATGTGAAGTGGCTATTTCTTTATATTTATCCACTACTCCGTGAAGGCTTCTCTGTTCTAC		
	ORF Start: ATG at 37		ORF Stop: TAA at 1039

	SEQ ID NO: 108	334 aa	MW at 37675.7kD
NOV26a, CG147284-01 Protein Sequence	MRTYRYFLLLFWVGQPYPTLSTPLSKRTSGFPAKKRALELSGNSKNELNRSKRSWMWNQFFLLEEYT GSDYQYVGKLHSNFTIQDNKDNTAGILTRKNGYNRHEMSTYLLPVVISDNDYPVQSSTGTVTVRVCA CDHHGNMQSCHAEALIHPTGLSTGALVAILLCIVILLVTVVLFAALRRQRKKEPLIISKEDIRDNIV SYNDEGGGEEDTQAFDIGTLRNPEAIEDNKLRRDIVPEALFLPRRTPTARDNTDVRDFINQRLKEND TDPTAPPYDSLATYAYEGTGSVADSLSSLESVTTDADQDYDYLSDWGPRFKKLADMYGGVDSDKDS		

Further analysis of the NOV26a protein yielded the following properties shown in Table 26B.

Table 26B. Protein Sequence Properties NOV26a	
PSort analysis:	0.7300 probability located in plasma membrane; 0.6400 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 22 and 23

A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 26C.

Table 26C. Geneseq Results for NOV26a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV26a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**

AAW13131	Partial human cadherin-6 - Homo sapiens, 414 aa. [US5597725-A, 28-JAN-1997]	79..334 159..414	256/256 (100%) 256/256 (100%)	e-148
AAW25659	Human cadherin-6 - Homo sapiens, 414 aa. [US5646250-A, 08-JUL-1997]	79..334 159..414	256/256 (100%) 256/256 (100%)	e-148
AAR43564	Human cadherin-6 - Homo sapiens, 391 aa. [WO9321302-A, 28-OCT-1993]	79..311 159..391	233/233 (100%) 233/233 (100%)	e-133
ABP47864	Human polypeptide SEQ ID NO 294 - Homo sapiens, 358 aa. [US2002042386-A1, 11-APR-2002]	77..334 101..358	212/258 (82%) 238/258 (92%)	e-125
AAU19644	Human novel extracellular matrix protein, Seq ID No 294 - Homo sapiens, 358 aa. [WO200155368-A1, 02-AUG-2001]	77..334 101..358	212/258 (82%) 238/258 (92%)	e-125

In a BLAST search of public sequence datbases, the NOV26a protein was found to have homology to the proteins shown in the BLASTP data in Table 26D.

Table 26D. Public BLASTP Results for NOV26a

Protein Accession Number	Protein/Organism/Length	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P55285	Cadherin-6 precursor (Kidney-cadherin) (K-cadherin) - Homo sapiens (Human), 790 aa.	79..334 535..790	256/256 (100%) 256/256 (100%)	e-147
P97326	Cadherin-6 precursor (Kidney-cadherin) (K-cadherin) - Mus musculus (Mouse), 790 aa.	79..334 535..790	246/256 (96%) 253/256 (98%)	e-143
P55280	Cadherin-6 precursor (Kidney-cadherin) (K-cadherin) - Rattus norvegicus (Rat), 789 aa.	79..334 535..789	239/256 (93%) 246/256 (95%)	e-136
Q90762	Cadherin-6 precursor (Cadherin-6B) (c-cad6B) - Gallus gallus (Chicken), 790 aa.	79..334 535..790	232/256 (90%) 243/256 (94%)	e-134

Q9DFS1	Cadherin-6 - Xenopus laevis (African clawed frog), 792 aa.	80..334 538..792	227/255 (89%) 240/255 (94%)	e-132

PFam analysis predicts that the NOV26a protein contains the domains shown in the Table 26E.

Table 26E. Domain Analysis of NOV26a

Pfam Domain	**NOV26a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Cadherin_C_term	182..328	108/156 (69%) 142/156 (91%)	6.6e-102

Example 27.

The NOV27 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis

	SEQ ID NO: 109	1082 bp	
NOV27a, CG147937-01 DNA Sequence	AAGTGGCTTCATTTCAGTGGCTGACTTCCAGAGAGCAAT**ATG**GCTGGTTCCCCAACATGCCTCACCC TCATCTATATCCTTTGGCAGCTCACAGGGTCAGCAGCCTCTGGACCCGTGAAAGAGCTGGTCGGTTC CGTTGGTGGGGCCGTGACTTTCCCCCTGAAGTCCAAAGTAAAGCAAGTTGACTCTATTGTCTGGACC TTCAACACAACCCCTCTTGTCACCATACAGCCAGAAGGGGGCACTATCATAGTGACCCAAAATCGTA ATAGGGAGAGAGTAGACTTCCCAGATGGAGGCTACTCCCTGAAGCTCAGCAAACTGAAGAAGAATGA CTCAGGGATCTACTATGTGGGGATATACAGCTCATCACTCCAGCAGCCCTCCACCCAGGAGTACGTG CTGCATGTCTACGAGCACCTGTCAAAGCCTAAAGTCACCATGGGTCTGCAGAGCAATAAGAATGGCA CCTGTGTGACCAATCTGACATGCTGCATGGAACATGGGGAAGAGGATGTGATTTATACCTGGAAGGC CCTGGGGCAAGCAGCCAATGAGTCCCATAATGGGTCCATCCTCCCCATCTCCTGGAGATGGGGAGAA AGTGATATGACCTTCATCTGCGTTGCCAGGAACCCTGTCAGCAGAAACTTCTCAAGCCCCATCCTTG CCAGGAAGCTCTGTGAAGGTGCTGCTGATGACCCAGATTCCTCCATGGTCCTCCTGTGTCTCCTGTT GGTGCCCCTCCTGCTCAGTCTCTTTGTACTGGGGCTATTTCTTTGGTTTCTGAAGAGAGAGAGACAA GAAGAGTACATTGAAGAGAAGAAGAGAGTGGACATTTGTCGGGAAACTCCTAACATATGCCCCCATT CTGGAGAGAACACAGAGTACGACACAATCCCTCACACTAATAGAACAATCCTAAAGGAAGATCCAGC AAATACGGTTTACTCCACTGTGGAAATACCGAAAAAGATGGAAAATCCCCACTCACTGCTCACGATG CCAGACACACCAAGGCTATTTGCCTATGAGAATGTTATC**TAG**ACAGCAGTGCACTCCCCTAAGTCTC TGCTCAAAAA		
	ORF Start: ATG at 40		ORF Stop: TAG at 1045

	SEQ ID NO: 110	335 aa	MW at 37420.5kD
NOV27a, CG147937-01 Protein Sequence	MAGSPTCLTLIYILWQLTGSAASGPVKELVGSVGGAVTFPLKSKVKQVDSIVWTFNTTPLVTIQPEG GTIIVTQNRNRERVDFPDGGYSLKLSKLKKNDSGIYYVGIYSSSLQQPSTQEYVLHVYEHLSKPKVT MGLQSNKNGTCVTNLTCCMEHGEEDVIYTWKALGQAANESHNGSILPISWRWGESDMTFICVARNPV SRNFSSPILARKLCEGAADDPDSSMVLLCLLLVPLLLSLFVLGLFLWFLKRERQEEYIEEKKRVDIC RETPNICPHSGENTEYDTIPHTNRTILKEDPANTVYSTVEIPKKMENPHSLLTMPDTPRLFAYENVI		

	SEQ ID NO: 111	1121 bp	
NOV27b,	AAGTGGCTTCATTTCAGTGGCTGACTTCCAGAGAGCAAT**ATG**GCTGGTTCCCCAACATGCCTCACCC TCATCTATATCCTTTGGCAGCTCACAGGGTCAGCAGCCTCTGGACCCGTGAAAGAGCTGGTCGGTTC		

CG147937-02 DNA Sequence	CGTTGGTGGGGCCGTGACTTTCCCCCTGAAGTCCAAAGTAAAGCAAGTTGACTCTATTGTCTGGACC TTCAACACAACCCCTCTTGTCACCATACAGCCAGAAGGGGGCACTATCATAGTGACCCAAAATCGTA ATAGGGAGAGAGTAGACTTCCCAGATGGAGGCTACTCCCTGAAGCTCAGCAAACTGAAGAAGAATGA CTCAGGGATCTACTATGTGGGGATATACAGCTCATCACTCCAGCAGCCCTCCACCCAGGAGTACGTG CTGCATGTCTACGAGCACCTGTCAAAGCCTAAAGTCACCATGGGTCTGCAGAGCAATAAGAATGGCA CCTGTGTGACCAATCTGACATGCTGCATGGAACATGGGGAAGAGGATGTGATTTATACCTGGAAGGC CCTGGGGCAAGCAGCCAATGAGTCCCATAATGGGTCCATCCTCCCCATCTCCTGGAGATGGGGAGAA AGTGATATGACCTTCATCTGCGTTGCCAGGAACCCTGTCAGCAGAAACTTCTCAAGCCCCATCCTTG CCAGGAAGCTCTGTGAAGGTGACTGCCTCTCCCCTCTCCACAGGAGACTCTGCCCAGGTGCTGCTGA TGACCCAGATTCCTCCATGGTCCTCCTGTGTCTCCTGTTGGTGCCCCTCCTGCTCAGTCTCTTTGTA CTGGGGCTATTTCTTTGGTTTCTGAAGAGAGAGAGACAAGAAGAGTACATTGAAGAGAAGAAGAGAG TGGACATTTGTCGGGAAACTCCTAACATATGCCCCCATTCTGGAGAGAACACAGAGTACGACACAAT CCCTCACACTAATAGAACAATCCTAAAGGAAGATCCAGCAAATACGGTTTACTCCACTGTGGAAATA CCGAAAAAGATGGAAAATCCCCACTCACTGCTCACGATGCCAGACACACCAAGGCTATTTGCCTATG AGAATGTTATCTAGACAGCAGTGCACTCCCCTAAGTCTCTGCTCAAAAA		
	ORF Start: ATG at 40		ORF Stop: TAG at 1084

	SEQ ID NO: 112	348 aa	MW at 38869.2kD
NOV27b, CG147937-02 Protein Sequence	MAGSPTCLTLIYILWQLTGSAASGPVKELVGSVGGAVTFPLKSKVKQVDSIVWTFNTTPLVTIQPEG GTIIVTQNRNRERVDFPDGGYSLKLSKLKKNDSGIYYVGIYSSSLQQPSTQEYVLHVYEHLSKPKVT MGLQSNKNGTCVTNLTCCMEHGEEDVIYTWKALGQAANESHNGSILPISWRWGESDMTFICVARNPV SRNFSSPILARKLCEGDCLSPLHRRLCPGAADDPDSSMVLLCLLLVPLLLSLFVLGLFLWFLKRERQ EEYIEEKKRVDICRETPNICPHSGENTEYDTIPHTNRTILKEDPANTVYSTVEIPKKMENPHSLLTM PDTPRLFAYENVI		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 27B.

Table 27B. Comparison of NOV27a against NOV27b.		
Protein Sequence	**NOV27a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV27b	1..335 1..348	290/348 (83%) 290/348 (83%)

Further analysis of the NOV27a protein yielded the following properties shown in Table 27C.

Table 27C. Protein Sequence Properties NOV27a	
PSort analysis:	0.4600 probability located in plasma membrane; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	Cleavage site between residues 23 and 24

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 27D.

Table 27D. Geneseq Results for NOV27a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV27a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB65224	Human PRO1138 (UNQ576) protein sequence SEQ ID NO:253 - Homo sapiens, 335 aa. [WO200073454-A1, 07-DEC-2000]	1..335 1..335	335/335 (100%) 335/335 (100%)	0.0
AAB87548	Human PRO1138 - Homo sapiens, 335 aa. [WO200116318-A2, 08-MAR-2001]	1..335 1..335	335/335 (100%) 335/335 (100%)	0.0
AAB47321	APEX-1 - Homo sapiens, 335 aa. [WO200146260-A2, 28-JUN-2001]	1..335 1..335	335/335 (100%) 335/335 (100%)	0.0
AAU29119	Human PRO polypeptide sequence #96 - Homo sapiens, 335 aa. [WO200168848-A2, 20-SEP-2001]	1..335 1..335	335/335 (100%) 335/335 (100%)	0.0
AAY66701	Membrane-bound protein PRO1138 - Homo sapiens, 335 aa. [WO9963088-A2, 09-DEC-1999]	1..335 1..335	335/335 (100%) 335/335 (100%)	0.0

In a BLAST search of public sequence datbases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27E.

Table 27E. Public BLASTP Results for NOV27a				
Protein Accession Number	**Protein/Organism/Length**	**NOV27a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9NQ25	BA404F10.4 (Novel LY9 (Lymphocyte antigen 9) like protein) (NK cell receptor) (Membrane protein FOAP-12) (CD2-like receptor activating cytotoxic cells) - Homo sapiens (Human), 335 aa.	1..335 1..335	335/335 (100%) 335/335 (100%)	0.0

Q9NY08	19A protein - Homo sapiens (Human), 335 aa.	1..335 1..335	334/335 (99%) 335/335 (99%)	0.0
Q9NY23	19A24 protein - Homo sapiens (Human), 328 aa.	1..316 1..281	273/316 (86%) 276/316 (86%)	e-152
AAH27867	19A24 protein - Homo sapiens (Human), 296 aa.	1..257 1..257	257/257 (100%) 257/257 (100%)	e-149
CAD39085	Hypothetical protein - Homo sapiens (Human), 228 aa.	120..335 12..228	212/217 (97%) 214/217 (97%)	e-123

PFam analysis predicts that the NOV27a protein contains the domains shown in the Table 27F.

Table 27F. Domain Analysis of NOV27a

Pfam Domain	**NOV27a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**

Example 28.

The NOV28 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 28A.

Table 28A. NOV28 Sequence Analysis

	SEQ ID NO: 113	561 bp	
NOV28a, CG148221-01 DNA Sequence	CTTGTGGCCCCCGGCTGCAGCCTCAGTGGC**ATG**GGGGTGAAGCGGAGCCTCCAGAGTGGGGGCATTC TGCTCAGCCTCGTGGCCAACGTCCTCATGGTGCTCTCCACGGCCACCAACTACTGGACCCGCCAACA AGAGGGCCACAGTGGCCTGTGGCAGGAATGCAACCACGGCATCTGCTCCAGCATCCCCTGCCAGAGT ACGCTGGCGGTGACTGTGGCGTGCATGGTGCTGGCGGTGGGTGTCGGCGTGGTGGGCATGGTGATGG GACTGCGGATTCGGTGCGACGAGGGCGAGTCGCTGCGGGGCCAGACCACGAGCGCCTTCCTCTTCCT CGGCGGACTGCTGCTGCTGACCGCCTTGATAGGCTACACCGTGAAGAATGCGTGGAAGAACAACGTC TTCTTCTCTTGGTCCTATTTTTCTGGGTGGCTGGCCTTACCCTTCTCAATTCTCGCGGGCTTCTGCT TTCTGCTGGCAGACATGATCATGCAGAGCACCGACGCCATCAGTGGATTCCCCGTGTGTCTG**TGA**CT GCAGCCTGCCTGGGGCAGAATAAAG		
	ORF Start: ATG at 31		ORF Stop: TGA at 532

	SEQ ID NO: 114	167 aa	MW at 17970.0kD
NOV28a, CG148221-01 Protein Sequence	MGVKRSLQSGGILLSLVANVLMVLSTATNYWTRQQEGHSGLWQECNHGICSSIPCQSTLAVTVACMV LAVGVGVVGMVMGLRIRCDEGESLRGQTTSAFLFLGGLLLLTALIGYTVKNAWKNNVFFSWSYFSGW LALPFSILAGFCFLLADMIMQSTDAISGFPVCL		

	SEQ ID NO: 115	561 bp	
NOV28b,	CTTGTGGCCCCCGGCTGCAGCCTCAGTGGC**ATG**GGGGTGAAGCGGAGCCTCCAGAGTGGGGGCATTC		

CG148221-02 DNA Sequence	TGCTCAGCCTCGTGGCCAACGTCCTCATGGTGCTCTCCACGGCCACCAACTACTGGACCCGCCAACA AGAGGGCCACAGTGGCCTGTGGCAGGAATGCAACCACGGCATCTGCTCCAGCATCCCCTGCCAGAGT ACGCTGGCGGTGACTGTGGCGTGCATGGTGCTGGCGGTGGGTGTCGGCGTGGTGGGCATGGTGATGG GACTGCGGATTCGGTGCGACGAGGGCGAGTCGCTGCGGGGCCAGACCACGAGCGCCTTCCTCTTCCT CGGCGGACTGCTGCTGCTGACCGCCTTGATAGGCTACACCGTGAAGAATGCGTGGAAGAACAACGTC TTCTTCTCTTGGTCCTATTTTTCTGGGTGGCTGGCCTTACCCTTCTCAATTCTCGCGGGCTTCTGCT TTCTGCTGGCAGACATGATCATGCAGAGCACCGACGCCATCAGTGGATTCCCCGTGTGTCTGTGACT GCAGCCTGCCTGGGGCAGAATAAAG		
	ORF Start: ATG at 31		ORF Stop: TGA at 532

	SEQ ID NO: 116	167 aa	MW at 17970.0kD
NOV28b, CG148221-02 Protein Sequence	MGVKRSLQSGGILLSLVANVLMVLSTATNYWTRQQEGHSGLWQECNHGICSSIPCQSTLAVTVACMV LAVGVGVVGMVMGLRIRCDEGESLRGQTTSAFLFLGGLLLLTALIGYTVKNAWKNNVFFSWSYFSGW LALPFSILAGFCFLLADMIMQSTDAISGFPVCL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 28B.

Table 28B. Comparison of NOV28a against NOV28b.		
Protein Sequence	**NOV28a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV28b	1..167 1..167	134/167 (80%) 134/167 (80%)

Further analysis of the NOV28a protein yielded the following properties shown in Table 28C.

Table 28C. Protein Sequence Properties NOV28a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6760 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 28 and 29

A search of the NOV28a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 28D.

Table 28D. Geneseq Results for NOV28a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV28a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**

AAU74822	Human REPTR 5 protein - Homo sapiens, 173 aa. [WO200198354-A2, 27-DEC-2001]	6..150 3..160	42/158 (26%) 71/158 (44%)	2e-07
AAR30057	Rat PMP - Rattus rattus, 160 aa. [WO9221694-A, 10-DEC-1992]	10..149 7..155	37/149 (24%) 66/149 (43%)	1e-06
AAB48599	Mouse PMP-22 - Mus sp, 161 aa. [US6150136-A, 21-NOV-2000]	10..149 7..155	38/149 (25%) 66/149 (43%)	3e-06
AAR30058	Mouse PMP - Mus musculus, 160 aa. [WO9221694-A, 10-DEC-1992]	10..149 7..155	38/149 (25%) 66/149 (43%)	3e-06
AAR30059	Bovine PMP - Bos taurus, 160 aa. [WO9221694-A, 10-DEC-1992]	10..149 7..155	37/150 (24%) 69/150 (45%)	2e-05

In a BLAST search of public sequence datbases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28E.

Table 28E. Public BLASTP Results for NOV28a				
Protein Accession Number	**Protein/Organism/Length**	**NOV28a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAH29518	Similar to RIKEN cDNA 1700071E18 gene - Homo sapiens (Human), 167 aa.	1..167 1..167	166/167 (99%) 167/167 (99%)	2e-93
Q9D9H2	1700071E18Rik protein - Mus musculus (Mouse), 167 aa.	1..167 1..167	117/167 (70%) 136/167 (81%)	2e-65
P54825	Lens fiber membrane intrinsic protein (MP17) (MP18) (MP19) (MP20) - Rattus norvegicus (Rat), 173 aa.	6..150 3..160	46/159 (28%) 73/159 (44%)	4e-08
P56563	Lens fiber membrane intrinsic protein (MP17) (MP18) (MP19) (MP20) - Mus musculus (Mouse), 173 aa.	6..150 3..160	46/159 (28%) 73/159 (44%)	4e-08
P20274	Lens fiber membrane intrinsic protein (MP18) (MP19) (MP21) (MP23) - Bos taurus (Bovine), 173 aa.	6..150 3..160	45/159 (28%) 72/159 (44%)	6e-07

PFam analysis predicts that the NOV28a protein contains the domains shown in the Table 28F.

Table 28F. Domain Analysis of NOV28a			
Pfam Domain	**NOV28a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
PMP22_Claudin	5..147	36/188 (19%) 108/188 (57%)	1.5e-05

Example 29.

The NOV29 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 29A.

Table 29A. NOV29 Sequence Analysis			
	SEQ ID NO: 117	2603 bp	
NOV29a, CG148476-01 DNA Sequence	CGTGGGCCGGGGTCGCGCAGCGGGCTGTGGGCGCGCCCGGAGGAGCGACCGCCGCAGTTCTCGAGCT CCAGCTGCATTCCCTCCGCGTCCGCCCCACGCTTCTCCCGCTCCGGGCCCCGCA**ATG**GCCCAGGCAG TGTGGTCGCGCCTCGGCCGCATCCTCTGGCTTGCCTGCCTCCTGCCCTGGGCCCCGGCAGGGGTGGC CGCAGGCCTGTATGAACTCAATCTCACCACCGATAGCCCTGCCACCACGGGAGCGGTGGTGACCATC TCGGCCAGCCTGGTGGCCAAGGACAACGGCAGCCTGGCCCTGCCCGCTGACGCCCACCTCTACCGCT TCCACTGGATCCACACCCCGCTGGTGCTTACTGGCAAGATGGAGAAGGGTCTCAGCTCCACCATCCG TGTGGTCGGCCACGTGCCCGGGGAATTCCCGGTCTCTGTCTGGGTCACTGCCGCTGACTGCTGGATG TGCCAGCCTGTGGCCAGGGGCTTTGTGGTCCTCCCCATCACAGAGTTCCTCGTGGGGGACCTTGTTG TCACCCAGAACACTTCCCTACCCTGGCCCAGCTCCTATCTCACTAAGACCGTCCTGAAAGTCTCCTT CCTCCTCCACGACCCGAGCAACTTCCTCAAGACCGCCTTGTTTCTCTACAGCTGGGACTTCGGGGAC GGGACCCAGATGGTGACTGAAGACTCCGTGGTCTATTATAACTATTCCATCATCCGGACCTTCACCG TGAAGCTCAAAGTGGTGGCGGAGTGGGAAGAGGTGGAGCCGGATGCCACGAGGGCTGTGAAGCAGAA GACCGGGGACTTCTCCGCCTCGCTGAAGCTGCAGGAAACCCTTCGAGGCATCCAAGTGTTGGGGCCC ACCCTAATTCAGACCTTCCAAAAGATGACCGTGACCTTGAACTTCCTGGGGAGCCCTCCTCTGACTG TGTGCTGGCGTCTCAAGCCTGAGTGCCTCCCGCTGGAGGAAGGGGAGTGCCACCCTGTGTCCGTGGC CAGCACAGCGTACAACCTGACCCACACCTTCAGGGACCCTGGGGACTACTGCTTCAGCATCCGGGCC GAGAATATCATCAGCAAGACACATCAGTACCACAAGATCCAGGTGTGGCCCTCCAGAATCCAGCCGG CTGTCTTTGCTTTCCCATGTGCTACACTTATCACTGTGATGTTGGCCTTCATCATGTACATGACCCT GCGGAATGCCACTCAGCAAAAGGACATGGTGGAGGTGGCTGATTTTGACTTTTCCCCCATGTCTGAC AAGAACCCGGAGCCACCCTCTGGGGTCAGGTGCTGCTGCCAGATGTGCTGTGGGCCTTTCTTGCTGG AGACTCCATCTGAGTACCTGGAAATTGTTCGTGAGAACCACGGGCTGCTCCCGCCCCTCTATAAGTC TGTCAAAACTTACACCGTG**TGA**GCACTCCCCCTCCCCACCCCATCTCAGTGTTAACTGACTGCTGAC TTGGAGTTTCCAGCAGGGTGGTGTGCACCACTGACCAGGAGGGGTTCATTTGCGTGGGGCTGTTGGC CTGGATCATCCATCCATCTGTACAGTTCAGCCACTGCCACAAGCCCCTCCCTCTCTGTCACCCCTGA CCCCAGCCATTCACCCATCTGTACAGTCCAGCCACTGACATAAGCCCCACTCGGTTACCACCCCCTT GACCCCCTACCTTTGAAGAGGCTTCGTGCAGGACTTTGATGCTTGGGGTGTTCCGTGTTGACTCCTA GGTGGGCCTGGCTGCCCACTGCCCATTCCTCTCATATTGGCACATCTGCTGTCCATTGGGGGTTCTC AGTTTCCTCCCCCAGACAGCCCTACCTGTGCCAGAGAGCTAGAAAGAAGGTCATAAAGGGTTAAAAA TCCATAACTAAAGGTTGTACACATAGATGGGCACACTCACAGAGAGAAGTGTGCATGTACACACACC ACACACACACACACACACACACACACACAGAAATATAAACACATGCGTCACATGGGCATTTCAGATGAT CAGCTCTGTATCTGGTTAAGTCGGTTGCTGGGATGCACCCTGCACTAGAGCTGAAAGGAAATTTGAC CTCCAAGCAGCCCTGACAGGTTCTGGGCCCGGGCCCTCCCTTTGTGCTTTGTCTCTGCAGTTCTTGC GCCCTTTATAAGGCCATCCTAGTCCCTGCTGGCTGGCAGGGGCCTGGATGGGGGGCAGGACTAATAC TGAGTGATTGCAGAGTGCTTTATAAATATCACCTTATTTTATCGAAACCCATCTGTGAAACTTTCAC TGAGGAAAAGGCCTTGCAGCGGTAGAAGAGGTTGAGTCAAGGCCGGGCGCGGTGGCTCACGCCTGTA ATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCACGAGATCAGGAGATCGAGACCACCCTGGCTA ACACGGTGAAACCCCGTCTCTACTAAAAAAATACAAAAAGTTAGCCGGGCGTGGTGGTGGGTGCCTG TAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGTGCGAACCCGGGAGGCGGAGCTTGCAGTG AGCCCAGATGGCGCCACTGCACTCCAGCCTGAGTGACAGAGCGAGACTCTGTCTCCA		

	ORF Start: ATG at 122		ORF Stop: TGA at 1427

	SEQ ID NO: 118	435 aa	MW at 48328.6kD
NOV29a, CG148476-01 Protein Sequence	MAQAVWSRLGRILWLACLLPWAPAGVAAGLYELNLTTDSPATTGAVVTISASLVAKDNGSLALPADA HLYRFHWIHTPLVLTGKMEKGLSSTIRVVGHVPGEFPVSVWVTAADCWMCQPVARGFVVLPITEFLV GDLVVTQNTSLPWPSSYLTKTVLKVSFLLHDPSNFLKTALFLYSWDFGDGTQMVTEDSVVYYNYSII GTFTVKLKVVAEWEEVEPDATRAVKQKTGDFSASLKLQETLRGIQVLGPTLIQTFQKMTVTLNFLGS PPLTVCWRLKPECLPLEEGECHPVSVASTAYNLTHTFRDPGDYCFSIRAENIISKTHQYHKIQVWPS RIQPAVFAFPCATLITVMLAFIMYMTLRNATQQKDMVEVADFDFSPMSDKNPEPPSGVRCCCQMCCG PFLLETPSEYLEIVRENHGLLPPLYKSVKTYTV		

Further analysis of the NOV29a protein yielded the following properties shown in Table 29B.

Table 29B. Protein Sequence Properties NOV29a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 25 and 26

A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 29C.

Table 29C. Geneseq Results for NOV29a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV29a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB65220	Human PRO1383 (UNQ719) protein sequence SEQ ID NO:241 - Homo sapiens, 423 aa. [WO200073454-A1, 07-DEC-2000]	1..435 1..423	423/435 (97%) 423/435 (97%)	0.0
AAM25558	Human protein sequence SEQ ID NO:1073 - Homo sapiens, 468 aa. [WO200153455-A2, 26-JUL-2001]	1..435 46..468	423/435 (97%) 423/435 (97%)	0.0
AAU29113	Human PRO polypeptide sequence #90 - Homo sapiens, 423 aa. [WO200168848-A2, 20-SEP-2001]	1..435 1..423	423/435 (97%) 423/435 (97%)	0.0

AAY66697	Membrane-bound protein PRO1383 - Homo sapiens, 423 aa. [WO9963088-A2, 09-DEC-1999]	1..435 1..423	423/435 (97%) 423/435 (97%)	0.0
ABG43580	Human peptide encoded by genome-derived single exon probe SEQ ID 33245 - Homo sapiens, 55 aa. [WO200186003-A2, 15-NOV-2001]	185..239 1..55	55/55 (100%) 55/55 (100%)	7e-24

In a BLAST search of public sequence datbases, the NOV29a protein was found to have homology to the proteins shown in the BLASTP data in Table 29D.

Table 29D. Public BLASTP Results for NOV29a

Protein Accession Number	Protein/Organism/Length	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD39014	Hypothetical protein - Homo sapiens (Human), 435 aa.	1..435 1..435	435/435 (100%) 435/435 (100%)	0.0
AAH30793	Similar to QNR-71 protein - Homo sapiens (Human), 423 aa.	1..435 1..423	423/435 (97%) 423/435 (97%)	0.0
CAD38628	Hypothetical protein --Homo sapiens (Human), 397 aa (fragment).	27..435 1..397	396/409 (96%) 396/409 (96%)	0.0
AAM31285	Surface layer protein B - Methanosarcina mazei (Methanosarcina frisia), 879 aa.	177..339 331..476	40/166 (24%) 64/166 (38%)	1e-04
AAH32783	Similar to glycoprotein (transmembrane) nmb - Homo sapiens (Human), 572 aa.	150..212 254..317	23/64 (35%) 36/64 (55%)	0.001

PFam analysis predicts that the NOV29a protein contains the domains shown in the Table 29E.

Table 29E. Domain Analysis of NOV29a

Pfam Domain	NOV29a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 30.

The NOV30 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 30A.

Table 30A. NOV30 Sequence Analysis			
	SEQ ID NO: 119	3273 bp	
NOV30a, CG148818-01 DNA Sequence	CTCCCGGAGATGCCCCGCGGCAGCCGCGCTCGGGGCTCTAAGAGAAAAAGGAGTTGGAATACAGAAT GCCCATCCTTTCCAGGAGAAAGACCACTGCAGGTCAGAAGAGCAGGTCTCAGGACAGCAGGGGCAGC TGCCTCTCTCTCTGAAGCATGGCTCAGGTGTGGAGAAGGGTTTCAGAACACTTCTGGGAATCCGTCA TTAACAGCTGAAGAGAAGACGATTACAGAAAAGCACCTTGAATTATGCCCTAGACCCAAGCAAGAAA CCACCACATCTAAAAGCACCAGTGGGCTTACAGACATAACATGGAGCTCCAGTGGAAGTGATTTGTC GGATGAAGATAAGACACTTTCTCAGTTACAGAGAGATGAATTACAGTTTATCGACTGGGAGATTGAC AGTGACAGGGCAGAGGCTAGTGACTGTGATGAATTTGAAGATGACGAGGGTGCTGTGGAAATCTCAG ACTGTGCTTCTTGTGCAAGTAATCAGTCTTTGACAAGTGATGAGAAGCTGTCGGAGCTTCCCAAGCC AAGTTCTATAGAAATTTTAGAGTATTCATCAGATAGTGAAAAAGAAGATGATTTGGAAAATGTCCTA CTCATTGATTCAGAATCCCCTCACAAATACCACGTGCAGTTTGCATCGGATGCAAGACAGATTATGG AGAGACTGATAGATCCAAGGACAAAATCAACAGAGACCATTTTGCATACACCTCAGAAACCCACAGC TAAGTTTCCCAGGACTCCAGAAAATTCAGCAAAGAAGAAGCTTTTAAGAGGTGGACTAGCAGAAAGA CTAAATGGACTGCAGAATCGAGAGAGATCTGCTATTTCTTTGTGGAGACATCAATGTATTTCTTACC AAAAGACACTTTCAGGTAGAAAATCTGGTGTATTAACTGTGAAAATTTTAGAGCTGCATGAGGAATG TGCCATGCAAGTTGCCATGTGTGAGCAGTTATTGGGGTCACCAGCCACCAGCTCCTCCCAAAGTGTG GCTCCCAGGCCTGGAGCTGGCCTGAAAGTTCTCTTCACCAAGGAGACTGCAGGCTACCTCAGGGGCC GTCCCCAGGACACTGTCCGGATCTTCCCTCCCTGGCAAAAACTGATTATTCCAAGTGGAAGTTGCCC TGTTATTCTGAATACTTACTTTTGTGAGAAAGTTGTTGCCAAAGAAGATTCAGAAAAAACTTGTGAA GTGTACTGTCCGGACATACCCCTTCCAAGAAGAAGCATCTCTTTGGCCCAGATGTTTGTAATTAAGG GTCTAACAAATAATTCACCTGAAATCCAGGTTGTGTGTAGTGGTGTAGCCACTACAGGGACAGCCTG GACCCATGGGCACAAAGAAGCAAAACAGCGCATCCCAACCAGCACTCCCCTGAGGGATTCTCTCCTG GATGTGGTGGAAAGCCAGGGAGCTGCCTCGTGGCCAGGAGCTGGAGTCCGAGTGGTGGTGCAAAGAG TGTATTCTCTTCCCAGCAGAGACAGCACCAGGGGTCAGCAGGGGGCCAGCTCAGGACACACAGACCC AGCTGGAACTCGAGCCTGCCTTCTGGTACAAGATGCCTGTGGAATGTTCGGTGAAGTGCACTTGGAG TTCACCATGTCGAAGGCAAGACAGTTGGAAGGGAAGTCTTGCAGCCTGGTGGGAATGAAGGTTCTAC AGAAAGTCACCAGAGGAAGGACAGCGGGGATTTTCAGTTTGATTGACACCCTGTGGCCCCCAGCGAT ACCTCTGAAAACACCTGGCCGCGACCAGCCCTGTGAAGAGATAAAAACTCATCTGCCTCCTCCAGCC TTGTGTTACATCCTCACAGCTCATCCAAATCTGGACAAATTGATATAATTGACGAAGACCCCATTT ATAAGCTTTACCAGCCTCCAGTTACCCGCTGCTTAAGAGACATTCTCCAGATGAATGATCTTGGTAC CCGTTGCAGTTTCTATGCCACGGTGATTTACCAAAAACCACAGCTGAAGAGTCTGCTGCTTCTGGAG CAAAGGGAGATCTGGCTGCTAGTGACCGATGTCACTCTGCAAACGAAGGAGGAGAGAGACCCCAGGC TCCCCAAAACCCTGCTGGTCTATGTGGCCCCCTTGTGTGTGCTGGGCTCTGAAGTCCTGGAGGCACT CGCTGGGGCTGCCCCTCACAGCCTCTTCTTCAAGGACGCTCTCCGTGACCAGGGTCGGATTGTTTGT GCTGAACGAACTGTCCTCTTGCTTCAGAAGCCCCTTTTGAGTGTGGTCTCTGGTGCAAGTTCCTGTG AGCTGCCTGGCCCGGTGATGCTCGACAGCCTGGACTCTGCAACACCTGTCAACTCCATCTGCAGTGT TCAAGGCACTGTGGTTGGCGTGGACGAGAGCACTGCTTTCTCATGGCCTGTGTGTGACATGTGTGGC AACGGGAGATTGGAACAGAGGCCGGAAGACAGAGGCGCCTTTTCCTGTGGGGACTGCTCCCGGGTGG TCACATCTCCTGTTCTCAAGAGGCACCTGCAGGTCTTCCTGGACTGCCGCTCAAGACCGCAGTGCAG AGTGAAGGTCAAGGTAGGAGCCAGGCCAGAGCACGCACGCACTCCTAGCTCACTCCAACATAGCGAA GCTGTTGCAGCGCAGCATTTCCTCCCTGCTGAGGTTTGCCGCCGGTGAAGATGGGAGCTACGAAGTG AAGAGTGTCCTCGGAAAGGAAGTGGGGTTGTTAAATTGTTTTGTCCAGTCCGTAACCGCCCACCCGA CCAGCTGCATTGGATTGGAGGAAATCGAGCCTCTGAGTGCAGGAGGGGCCTCTGCAGAACACTAGCG GTTGCCGCAGGATCTGTGAACTTTGCAATGTGGCTGCAAGGGTGGTGGTGGTGGTGGTGATTTGGGG TAGTTATTTGTTAACTATGGACACAGTGAACGTAGTTTACGATCTTGAAATGAAACTTAGATTTTTC TGGGGAAATGTTCAGATACAGTTTTGTGAACTGTAAATCAAAATACCTTTTTCTACAGTTTATCTTT TATTTTCTGCAAATTTAGGAACATATTTACTCGTTTTCACATTGAATCTTAAGTTTAAGCTCTTCAT TTGGTATTTAGGCAATATATGAGAAAAAAATTTTTTTTGTTCATTTGTAATTTTAACAAGTTGAACA TTTTACCATGATTGAACATGTTTTTATTACAGTATTTAACATTCCCCCAAAGAATACCCTGCAAAGT GTAAACCTTTGTCCCATACTGTGATATTACTGTTCTGCTACAATAAATGTCAAACCT		
	ORF Start: ATG at 10		ORF Stop: TGA at 2659

	SEQ ID NO: 120	883 aa	MW at 97134.4kD
NOV30a, CG148818-01 Protein Sequence	MPRGSRARGSKRKRSWNTECPSFPGERPLQVRRAGLRTAGAAASLSEAWLRCGEGFQNTSGNPSLTA EEKTITEKHLELCPRPKQETTTSKSTSGLTDITWSSSGSDLSDEDKTLSQLQRDELQFIDWEIDSDR AEASDCDEFEDDEGAVEISDCASCASNQSLTSDEKLSELPKPSSIEILEYSSDSEKEDDLENVLLID SESPHKYHVQFASDARQIMERLIDPRTKSTETILHTPQKPTAKFPRTPENSAKKKLLRGGLAERLNG LQNRERSAISLWRHQCISYQKTLSGRKSGVLTVKILELHEECAMQVAMCEQLLGSPATSSSQSVAPR PGAGLKVLFTKETAGYLRGRPQDTVRIFPPWQKLIIPSGSCPVILNTYFCEKVVAKEDSEKTCEVYC PDIPLPRRSISLAQMFVIKGLTNNSPEIQVVCSGVATTGTAWTHGHKEAKQRIPTSTPLRDSLLDVV ESQGAASWPGAGVRVVVQRVYSLPSRDSTRGQQGASSGHTDPAGTRACLLVQDACGMFGEVHLEFTM SKARQLEGKSCSLVGMKVLQKVTRGRTAGIFSLIDTLWPPAIPLKTPGRDQPCEEIKTHLPPPALCY ILTAHPNLGQIDIIDEDPIYKLYQPPVTRCLRDILQMNDLGTRCSFYATVIYQKPQLKSLLLLEQRE IWLLVTDVTLQTKEERDPRLPKTLLVYVAPLCVLGSEVLEALAGAAPHSLFFKDALRDQGRIVCAER TVLLLQKPLLSVVSGASSCELPGPVMLDSLDSATPVNSICSVQGTVVGVDESTAFSWPVCDMCGNGR LEQRPEDRGAFSCGDCSRVVTSPVLKRHLQVFLDCRSRPQCRVKVKVGARPEHARTPSSLQHSEAVA AQHFLPAEVCRR		

Further analysis of the NOV30a protein yielded the following properties shown in Table 30B.

Table 30B. Protein Sequence Properties NOV30a	
PSort analysis:	0.4400 probability located in plasma membrane; 0.4284 probability located in mitochondrial inner membrane; 0.2397 probability located in mitochondrial matrix space; 0.2397 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV30a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 30C.

Table 30C. Geneseq Results for NOV30a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG76160	Human colon cancer antigen protein SEQ ID NO:6924 - Homo sapiens, 43 aa. [WO200122920-A2, 05-APR-2001]	175..206 12..43	29/32 (90%) 30/32 (93%)	5e-08
ABB60641	Drosophila melanogaster polypeptide SEQ ID NO 8715 - Drosophila melanogaster, 476 aa. [WO200171042-A2, 27-SEP-2001]	59..255 284..462	51/199 (25%) 85/199 (42%)	0.004
ABP39618	Staphylococcus epidermidis	17..237	52/224 (23%)	0.017

	ORF amino acid sequence SEQ ID NO:4463 - Staphylococcus epidermidis, 2137 aa. [US6380370-B1, 30-APR-2002]	1831..2048	88/224 (39%)	
AAB30809	Amino acid sequence of a prion-like amyloidogenic protein - Saccharomyces cerevisiae, 414 aa. [WO200075324-A2, 14-DEC-2000]	64..196 30..154	32/133 (24%) 61/133 (45%)	0.030
AAW10529	Saccharomyces cerevisiae nucleolin like protein, NOL1 - Saccharomyces cerevisiae (S288C), 414 aa. [US5470971-A, 28-NOV-1995]	64..196 30..154	32/133 (24%) 61/133 (45%)	0.030

In a BLAST search of public sequence datbases, the NOV30a protein was found to have homology to the proteins shown in the BLASTP data in Table 30D.

Table 30D. Public BLASTP Results for NOV30a

Protein Accession Number	Protein/Organism/Length	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q14159	KIAA0146 protein - Homo sapiens (Human), 918 aa (fragment).	1..851 4..854	850/851 (99%) 851/851 (99%)	0.0
Q8R305	Hypothetical 43.0 kDa protein - Mus musculus (Mouse), 393 aa.	527..851 6..325	223/325 (68%) 258/325 (78%)	e-125
Q96BI5	Hypothetical 23.1 kDa protein - Homo sapiens (Human), 218 aa (fragment).	701..851 4..154	150/151 (99%) 151/151 (99%)	1e-82
P97399	Dentin sialophosphoprotein precursor (Dentin matrix protein-3) (DMP- 3) [Contains: Dentin phosphoprotein (Dentin phosphophoryn) (DPP) Dentin sialoprotein (DSP)] - Mus musculus (Mouse), 934 aa.	86..196 581..692	29/112 (25%) 50/112 (43%)	0.029

Q01538	Myelin transcription factor 1 (MYT1) (MYTI) (Proteolipid protein binding protein) (PLPB1) - Homo sapiens (Human), 1121 aa.	67..193 221..355	32/136 (23%) 65/136 (47%)	0.051

PFam analysis predicts that the NOV30a protein contains the domains shown in the Table 30E.

Table 30E. Domain Analysis of NOV30a			
Pfam Domain	**NOV30a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
zf-B_box	792..837	11/49 (22%) 32/49 (65%)	0.48

Example 31.

The NOV31 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 31A.

Table 31A. NOV31 Sequence Analysis			
	SEQ ID NO: 121	447 bp	
NOV31a, CG149332-01 DNA Sequence	ACCATGAACCACACTGTCCAAACCCTCTTCACTCCTGCCAACACCGGCCGCTCCACCAACCATGAGA TGCTCAAGGAGAAGCATGAGGTGGCTGTGCTGGGGGCACCCCACAACCCTGTGCCTCCAGCGTTCAC CATGATCCACATCTGCAGTGAGACCTCCGTGCCCGACCATGTCGTCTGGTCCCTATTCAACACCCTC TTCAAGAATTCCTGCTGCCCGGACTTCATAGCATTCATCTACTCTGTGAAGTCTAGGGACAGGAAGA TGGTTGGTGACCTGACTGGGGCCCAGGCCTGTGTCTCCACTGCCAAGTGCCTGAACATCTGGGCCCT GGCTCTGGGCATCCTCCTGACCATTCTGCTCATCATCATCTCAGTGCTGATCTTCCAAGTCTCTCGA TAGAACAGGAGACAGCATCCGGGCCAGGAGCTCTGCCCACAACCT		
	ORF Start: ATG at 4		ORF Stop: TAG at 403

	SEQ ID NO: 122	133 aa	MW at 14678.1kD
NOV31a, CG149332-01 Protein Sequence	MNHTVQTLFTPANTGRSTNHEMLKEKHEVAVLGAPHNPVPPAFTMIHICSETSVPDHVVWSLFNTLF KNSCCPDFIAFIYSVKSRDRKMVGDLTGAQACVSTAKCLNIWALALGILLTILLIIISVLIFQVSR		

Further analysis of the NOV31a protein yielded the following properties shown in Table 31B.

Table 31B. Protein Sequence Properties NOV31a

PSort analysis:	0.7000 probability located in plasma membrane; 0.2000 probability located in endoplasmic reticulum (membrane); 0.1242 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV31a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 31C.

Table 31C. Geneseq Results for NOV31a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABP43105	Human ovarian antigen HVCBB19, SEQ ID NO:4237 - Homo sapiens, 143 aa. [WO200200677-A1, 03-JAN-2002]	1..130 11..140	104/130 (80%) 113/130 (86%)	3e-56
AAE13797	Human lung tumour-specific protein SALT-T8 - Homo sapiens, 133 aa. [WO200172295-A2, 04-OCT-2001]	1..130 1..130	104/130 (80%) 113/130 (86%)	3e-56
AAB44456	Human lung tumour-specific antigen encoded by cDNA #71 - Homo sapiens, 133 aa. [WO200060077-A2, 12-OCT-2000]	1..130 1..130	104/130 (80%) 113/130 (86%)	3e-56
AAY29544	Human lung tumour protein SALT-T8 predicted amino acid sequence - Homo sapiens, 133 aa. [WO9938973-A2, 05-AUG-1999]	1..130 1..130	104/130 (80%) 113/130 (86%)	3e-56
AAY93594	Protein encoded by I-8U gene from interferon-inducible gene family - Homo sapiens, 133 aa. [WO200035473-A2, 22-JUN-2000]	1..130 1..130	102/130 (78%) 112/130 (85%)	3e-55

In a BLAST search of public sequence datbases, the NOV31a protein was found to have homology to the proteins shown in the BLASTP data in Table 31D.

Table 31D. Public BLASTP Results for NOV31a				
Protein Accession Number	Protein/Organism/Length	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q01628	Interferon-induced transmembrane protein 3 (Interferon-inducible protein 1-8U) - Homo sapiens (Human), 133 aa.	1..130 1..130	104/130 (80%) 113/130 (86%)	7e-56
AAH22439	Interferon induced transmembrane protein 3 (1-8U) - Homo sapiens (Human), 133 aa.	1..130 1..130	103/130 (79%) 112/130 (85%)	6e-55
S17182	interferon-induced protein 1-8U - human, 133 aa.	1..130 1..130	102/130 (78%) 112/130 (85%)	8e-55
Q01629	Interferon-induced transmembrane protein 2 (Interferon-inducible protein 1-8D) - Homo sapiens (Human), 132 aa.	1..133 1..132	98/133 (73%) 110/133 (82%)	2e-51
Q95MQ3	Interferon-induced protein 1-8U - Bos taurus (Bovine), 146 aa.	1..124 1..124	78/124 (62%) 97/124 (77%)	9e-39

PFam analysis predicts that the NOV31a protein contains the domains shown in the Table 31E.

Table 31E. Domain Analysis of NOV31a			
Pfam Domain	NOV31a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 32.

The NOV32 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 32A.

Table 32A. NOV32 Sequence Analysis			
	SEQ ID NO: 123	702 bp	

NOV32a, CG149649-01 DNA Sequence	GGTGGTCAGGGCGCCATGGCGCTGTCCTGGCTGCAGCGCGTCGAGCTTGCGCTCTTTGCTGCCGCCT TCCTGTGCGGGGCCGTGGCGGCCGCGGCGATGACTCGGACCCAGGGCTCCTTCAGTGGTAGATGTCC CCTGTATGGTGTGGCCACCCTGAATGGCTCCTCCCTGGCCTTATCCCGTCCCTCAGCACCATCCCTG TGCTACTTTGTAGCTGGGGCCTCTGGCCTCTTGGCCCTCTACTGCCTCCTGCTTTTGCTCTTCTGGA TCTACAGCAGCTGCATCGAGGACTCCCACAGGGGTGCTATAGGGCTGCGCATTGCACTGGCCATCTC AGCTATAGCCGTCTTCCTGGTCTTGGTGTCTGCCTGTATCCTTCGATTTGGCACCAGGTCTCTCTGC AACTCCATCATCTCCTTGAACACTACAATTAGCTGTTCTGAAGCCCAGAAAATTCCATGGACACCCC CTGGAACTGCTCTGCAGTTTTACTCCAACCTACACAATGCTGAAACCTCTTCTTGGGTGAATTTGGT ATTGTGGTGTGTGGTCTTGGTGCTCCAGGTCGTGCAGTGGAAGTCTGAAGCCACCCCATACCGGCCT CTGGAGAGGGGTGACCCTGAGTGGAGCTCTGAGACAGATGCTCTCGTTGGGTCACGCCTTTCCCATT CCTGAAGAATAAGCGGAGTGCTTCCTGCAGCC		
	ORF Start: ATG at 16		ORF Stop: TGA at 673

	SEQ ID NO: 124	219 aa	MW at 23550.0kD
NOV32a, CG149649-01 Protein Sequence	MALSWLQRVELALFAAAFLCGAVAAAAMTRTQGSFSGRCPLYGVATLNGSSLALSRPSAPSLCYFVA GASGLLALYCLLLLLFWIYSSCIEDSHRGAIGLRIALAISAIAVFLVLVSACILRFGTRSLCNSIIS LNTTISCSEAQKIPWTPPGTALQFYSNLHNAETSSWVNLVLWCVVLVLQVVQWKSEATPYRPLERGD PEWSSETDALVGSRLSHS		

	SEQ ID NO: 125	708 bp	
NOV32b, CG149649-02 DNA Sequence	GTGCTGGAATTCGCCCTTCATGGCGCTGTCCTGGCTGCAGCGCGTCGAGCTTGCGCTCTTTGCTGCC GCCTTCCTGTGCGGGGCCGTGGCGGCCGCGGCGATGACTCGGACCCAGGGCTCCTTCAGTGGTAGAT GTCCCCTGTATGGTGTGGCCACCCTGAATGGCTCCTCCCTGGCCTTATCCCGTCCCTCAGCACCATC CCTGTGCTACTTTGTAGCTGGGGCCTCTGGCCTCTTGGCCCTCTACTGCCTCCTGCTTTTGCTCTTC TGGATCTACAGCAGCTGCATCGAGGACTCCCACAGAGGTGCTATAGGGCTGCGCATTGCACTGGCCA TCTCAGCTATAGCCGTCTTCCTGGTCTTGGTGTCTGCCTGTATCCTTCGATTTGGCACCAGGTCTCT CTGCAACTCCATCATCTCTTTGAACACTACAATTAGCTGTTCTGAAGCCCAGAAAATTCCATGGACA CCCCCTGGAACTGCTCTGCAGTTTTACTCCAACCTACACAATGCTGAAACCTCTTCTTGGGTGAATT TGGTATTGTGGTGTGTGGTCTTGGTGCTCCAGGTCGTGCAGTGGAAGTCTGAAGCCACCCCATACCG GCCTCTGGAGAGGGGTGACCCTGAGTGGAGCTCTGAGACAGATGCTCTCGTTGGGTCACGCCTTTCC CATTCCTGAAGAATAAGCGGAGTGCTAAGGGCGATTCC		
	ORF Start: ATG at 20		ORF Stop: TGA at 677

	SEQ ID NO: 126	219 aa	MW at 23550.0kD
NOV32b, CG149649-02 Protein Sequence	MALSWLQRVELALFAAAFLCGAVAAAAMTRTQGSFSGRCPLYGVATLNGSSLALSRPSAPSLCYFVA GASGLLALYCLLLLLFWIYSSCIEDSHRGAIGLRIALAISAIAVFLVLVSACILRFGTRSLCNSIIS LNTTISCSEAQKIPWTPPGTALQFYSNLHNAETSSWVNLVLWCVVLVLQVVQWKSEATPYRPLERGD PEWSSETDALVGSRLSHS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 32B.

Table 32B. Comparison of NOV32a against NOV32b.		
Protein Sequence	**NOV32a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV32b	1..219 1..219	160/219 (73%) 160/219 (73%)

Further analysis of the NOV32a protein yielded the following properties shown in Table 32C.

Table 32C. Protein Sequence Properties NOV32a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 25 and 26

A search of the NOV32a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 32D.

Table 32D. Geneseq Results for NOV32a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV32a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU12071	Human PHT1 variant protein from Caco-2 cells - Homo sapiens, 577 aa. [WO200192468-A2, 06-DEC-2001]	19..138 14..142	40/134 (29%) 54/134 (39%)	3.3
AAU12070	Human PHT1 variant protein from BeWo cells - Homo sapiens, 577 aa. [WO200192468-A2, 06-DEC-2001]	19..138 14..142	40/134 (29%) 54/134 (39%)	3.3
AAU12069	Human PHT1 protein splice variant - Homo sapiens, 295 aa. [WO200192468-A2, 06-DEC-2001]	19..138 14..142	40/134 (29%) 54/134 (39%)	3.3
AAU12068	Human PHT1 protein isolated from Caco-2 cells - Homo sapiens, 577 aa. [WO200192468-A2, 06-DEC-2001]	19..138 14..142	40/134 (29%) 54/134 (39%)	3.3
ABB91559	Herbicidally active polypeptide SEQ ID NO 770 - Arabidopsis thaliana, 763 aa. [WO200210210-A2, 07-FEB-2002]	32..125 603..689	20/94 (21%) 42/94 (44%)	5.6

In a BLAST search of public sequence datbases, the NOV32a protein was found to have homology to the proteins shown in the BLASTP data in Table 32E.

Table 32E. Public BLASTP Results for NOV32a				
Protein Accession Number	Protein/Organism/Length	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9CY24	1810059G22Rik protein - Mus musculus (Mouse), 219 aa.	1..219 1..219	177/219 (80%) 191/219 (86%)	e-100
Q9D8L7	1810059G22Rik protein - Mus musculus (Mouse), 219 aa.	1..219 1..219	176/219 (80%) 191/219 (86%)	e-100
Q9FLD9	Similarity to hedgehog-interacting protein - Arabidopsis thaliana (Mouse-ear cress), 677 aa.	114..182 611..676	24/69 (34%) 35/69 (49%)	0.60
O83823	Hypothetical protein TP0851 - Treponema pallidum, 724 aa.	96..143 280..326	20/48 (41%) 28/48 (57%)	1.4
Q9JVM8	Hypothetical protein NMA0774 - Neisseria meningitidis (serogroup A), 352 aa.	132..188 156..213	15/58 (25%) 30/58 (50%)	5.2

PFam analysis predicts that the NOV32a protein contains the domains shown in the Table 32F.

Table 32F. Domain Analysis of NOV32a			
Pfam Domain	NOV32a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 33.

The NOV33 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 33A.

Table 33A. NOV33 Sequence Analysis			
	SEQ ID NO: 127	2105 bp	
NOV33a, CG149680-01 DNA Sequence	AGGTGCAAAGCCTGGTGCCCCGAGCCCTGCGGAGCTCGGGGCCAGCATGGCCCCCACGCTGCAACAG GCGTACCGGAGGCGCTGGTGGATGGCCTGCACGGCTGTGCTGGAGAACCTCTTCTTCTCTGCTGTAC TCCTGGGCTGGGGCTCCCTGTTGATCATTCTGAAGAACGAGGGCTTCTATTCCAGCACGTGCCCAGC TGAGAGCAGCACCAACACCACCCAGGATGAGCAGCGCAGGTGGCCTTGCTTCACTGCGTCCTGCACC		

	CTCATGGCCCTGGCCTCCCGGGACGTGGAAGCTCTGTCTCCGTTGATATTCCTGGCGCTGTCCCTGA ATGGCTTTGGTGGCATCTGCCTAACGTTCACTTCACTCACGCTGCCCAACATGTTTGGGAACCTGCG CTCCACGTTAATGGCCCTCATGATTGGCTCTTACGCCTCTTCTGCCATTACGTTCCCAGGAATCAAG CTGATCTACGATGCCGGTGTGGCCTTCGTGGTCATCATGTTCACCTGGTCTGGCCTGGCCTGCCTTA TCTTTCTGAACTGCACCCTCAACTGGCCCATCGAAGCCTTTCCTGCCCCTGAGGAAGTCAATTACAC GAAGAAGATCAAGCTGAGTGGGCTGGCCCTGGACCACAAGGTGACAGGTGACCTCTTCTACACCCAT GTGACCACCATGGGCCAGAGGCTCAGCCAGAAGGCCCCCAGCCTGGAGGACGGTTCGGATGCCTTCA TGTCACCCCAGGATGTTCGGGGCACCTCAGAAAACCTTCCTGAGAGGTCTGTCCCCTTACGCAAGAG CCTCTGCTCCCCCACTTTCCTGTGGAGCCTCCTCACCATGGGCATGACCCAGCTGCGGATCATCTTC TACATGGCTGCTGTGAACAAGATGCTGGAGTACCTTGTGACTGGTGGCCAGGAGCATGAGACAAATG AACAGCAACAAAAGGTGGCAGAGACAGTTGGGTTCTACTCCTCCGTCTTCGGGGCCATGCAGCTGTT GTGCCTTCTCACCTGCCCCCTCATTGGCTACATCATGGACTGGCGGATCAAGGACTGCGTGGACGCC CCAACTCAGGGCACTGTCCTCGGAGATGCCAGGGACGGGGTTGCTACCAAATCCATCAGACCACGCT ACTGCAAGATCCAAAAGCTCACCAATGCCATCAGTGCCTTCACCCTGACCAACCTGCTGCTTGTGGG TTTTGGCATCACCTGTCTCATCAACAACTTACACCTCCAGTTTGTGACCTTTGTCCTGCACACCATT GTTCGAGGTTTCTTCCACTCAGCCTGTGGGAGTCTCTATGCTGCAGTGTTCCCATCCAACCACTTTG GGACGCTGACAGGCCTGCAGTCCCTCATCAGTGCTGTGTTCGCCTTGCTTCAGCAGCCACTTTTCAT GGCGATGGTGGGACCCCTGAAAGGAGAGCCCTTCTGGGTGAATCTGGGCCTCCTGCTATTCTCACTC CTGGGATTCCTGTTGCCTTCCTACCTCTTCTATTACCGTGCCCGGCTCCAGCAGGAGTACGCCGCCA ATGGGATGGGCCCACTGAAGGTGCTTAGCGGCTCTGAGGTGACCGCATAGACTTCTCAGACCAAGGG ACCTGGATGACAGGCAATCAAGGCCTGAGCAACCAAAAGGAGTGCCCCATATGGCTTTTCTACCTGT AACATGCACATAGAGCCATGGCCGTAGATTTATAAATACCAAGAGAAGTTCTATTTTTGTAAAGACT GCAAAAAGGAGGAAAAAAAAACCTTCAAAAACGCCCCCTAAGTCAACGCTCCATTGACTGAAGACAG TCCCTATCCTAGAGGGGTTGAGCTTTCTTCCTCCTTGGGTTGGAGGAGACCAGGGTGCCTCTTATCT CCTTCTAGCGGTCTGCCTCCTGGTACCTCTTGGGGGGATCGGCAAACAGGCTACCCCTGAGGTCCCA TGTGCCATGAGTGTGCACACATGCATGTGTCTGTGTATGTGTGAATGTGAGAGAGACACAGCCCTCC TTTCAGAAGGAAAGGGGCCTGAGGTGCCAGCTGTGTCCTGGGTTAGGGGTTGGGGGTCGGCCCCTTC CAGGGCCAGGAGGTCAGGTTCCTCAGCG		
	ORF Start: ATG at 47		ORF Stop: TAG at 1589

	SEQ ID NO: 128	514 aa	MW at 56699.6kD
NOV33a, CG149680-01 Protein Sequence	MAPTLQQAYRRRWWMACTAVLENLFFSAVLLGWGSLLIILKNEGFYSSTCPAESSTNTTQDEQRRWP CFTASCTLMALASRDVEALSPLIFLALSLNGFGGICLTFTSLTLPNMFGNLRSTLMALMIGSYASSA ITFPGIKLIYDAGVAFVVIMFTWSGLACLIFLNCTLNWPIEAFPAPEEVNYTKKIKLSGLALDHKVT GDLFYTHVTTMGQRLSQKAPSLEDGSDAFMSPQDVRGTSENLPERSVPLRKSLCSPTFLWSLLTMGM TQLRIIFYMAAVNKMLEYLVTGGQEHETNEQQQKVAETVGFYSSVFGAMQLLCLLTCPLIGYIMDWR IKDCVDAPTQGTVLGDARDGVATKSIRPRYCKIQKLTNAISAFTLTNLLLVGFGITCLINNLHLQFV TFVLHTIVRGFFHSACGSLYAAVFPSNHFGTLTGLQSLISAVFALLQQPLFMAMVGPLKGEPFWVNL GLLLFSLLGFLLPSYLFYYRARLQQEYAANGMGPLKVLSGSEVTA		

	SEQ ID NO: 129	2284 bp	
NOV33b, CG149680-02 DNA Sequence	AGGTGCAAAGCCTGGTGCCCCGAGCCCTGCGGAGCTCGGGCCCAGCATGGCCCCCACGCTGCAACAG GCGTACCGGAGGCGCTGGTGGATGGCCTGCACGGCTGTGCTGGAGAACCTCTTCTTCTCTGCTGTAC TCCTGGGCTGGGGCTCCCTGTTGATCATTCTGAAGAACGAGGGCTTCTATTCCAGCACGTGCCCAGC TGTTCCTGGTGTCATGTGCTGGGCCCTCCCTTCCCCCTCCTCAGCTGAGAGCAGCACCAACACCACC CAGGATGAGCAGCGCAGGTGGCCAGGCTGTGACCAGCAGGACGAGATGCTCAACCTGGGCTTCACCA TTGGTTCCTTCGTGCTCAGCGCCACCACCCTGCCACTGGGGATCCTCATGGACCGCTTTGGCCCCCG ACCCGTGCGGCTGGTTGGCAGTGCCTGCTTCACTGCGTCCTGCACCCTCATGGCCCTGGCCTCCCGG GACGTGGAAGCTCTGTCTCCGTTGATATTCCTGGCGCTGTCCCTGAATGGCTTTGGTGGCATCTGCC TAACGTTCACTTCACTCACGCTGCCCAACATGTTTGGGAACCTGCGCTCCACGTTAATGGCCCTCAT GATTGGCTCTTACGCCTCTTCTGCCATTACGTTCCCAGGAATCAAGCTGATCTACGATGCCGGTGTG GCCTTCGTGGTCATCATGTTCACCTGGTCTGGCCTGGCCTGCCTTATCTTTCTGAACTGCACCCTCA ACTGGCCCATCGAAGCCTTTCCTGCCCCTGAGGAAGTCAATTACACGAAGAAGATCAAGCTGAGTGG GCTGGCCCTGGACCACAAGGTGACAGGTGACCTCTTCTACACCCATGTGACCACCATGGGCCAGAGG CTCAGCCAGAAGGCCCCCAGCCTGGAGGACGGTTCGGATGCCTTCATGTCACCCCAGGATGTTCGGG GCACCTCAGAAAACCTTCCTGAGAGGTCTGTCCCCTTACGCAAGAGCCTCTGCTCCCCCACTTTCCT GTGGAGCCTCCTCACCATGGGCATGACCCAGCTGCGGATCATCTTCTACATGGCTGCTGTGAACAAG ATGCTGGAGTACCTTGTGACTGGTGGCCAGGAGCATGAGACAAATGAACAGCAACAAAAGGTGGCAG AGACAGTTGGGTTCTACTCCTCCGTCTTCGGGGCCATGCAGCTGTTGTGCCTTCTCACCTGCCCCCT CATTGGCTACATCATGGACTGGCGGATCAAGGACTGCGTGGACGCCCCAACTCAGGGCACTGTCCTC GGAGATGCCAGGGACGGGGTTGCTACCAAATCCATCAGACCACGCTACTGCAAGATCCAAAAGCTCA CCAATGCCATCAGTGCCTTCACCCTGACCAACCTGCTGCTTGTGGGTTTTGGCATCACCTGTCTCAT CAACAACTTACACCTCCAGTTTGTGACCTTTGTCCTGCACACCATTGTTCGAGGTTTCTTCCACTCA GCCTGTGGGAGTCTCTATGCTGCAGTGTTCCCATCCAACCACTTTGGGACGCTGACAGGCCTGCAGT CCCTCATCAGTGCTGTGTTCGCCTTGCTTCAGCAGCCACTTTTCATGGCGATGGTGGGACCCCTGAA AGGAGAGCCCTTCTGGGTGAATCTGGGCCTCCTGCTATTCTCACTCCTGGGATTCCTGTTGCCTTCC TACCTCTTCTATTACCGTGCCCGGCTCCAGCAGGAGTACGCCGCCAATGGGATGGGCCCACTGAAGG TGCTTAGCGGCTCTGAGGTGACCGCATAGACTTCTCAGACCAAGGGACCTGGATGACAGGCAATCAA GGCCTGAGCAACCAAAAGGAGTGCCCCATATGGCTTTTCTACCTGTAACATGCACATAGAGCCATGG		

	CCGTAGATTTATAAATACCAAGAGAAGTTCTATTTTTGTAAAGACTGCAAAAAGGAGGAAAAAAAAC CTTCAAAAACGCCCCCTAAGTCAACGCTCCATTGACTGAAGACAGTCCCTATCCTAGAGGGGTTGAG CTTTCTTCCTCCTTGGGTTGGAGGAGACCAGGGTGCCTCTTATCTCCTTCTAGCGGTCTGCCTCCTG GTACCTCTTGGGGGGATCGGCAAACAGGCTACCCCTGAGGTCCCATGTGCCATGAGTGTGCACACAT GCATGTGTCTGTGTATGTGTGAATGTGAGAGAGACACAGCCCTCCTTTCAGAAGGAAAGGGGCCTGA GGTGCCAGCTGTGTCCTGGGTTAGGGGTTGGGGGTCGGCCCCTTCCAGGGCCAGGAGGTCAGGTTCC TCAGCG		
	ORF Start: ATG at 47		ORF Stop: TAG at 1769

	SEQ ID NO: 130	574 aa	MW at 62959.8kD
NOV33b, CG149680-02 Protein Sequence	MAPTLQQAYRRRWWMACTAVLENLFFSAVLLGWGSLLIILKNEGFYSSTCPAVPGVMCWALPSPSSA ESSTNTTQDEQRRWPGCDQQDEMLNLGFTIGSFVLSATTLPLGILMDRFGPRPVRLVGSACFTASCT LMALASRDVEALSPLIFLALSLNGFGGICLTFTSLTLPNMFGNLRSTLMALMIGSYASSAITFPGIK LIYDAGVAFVVIMFTWSGLACLIFLNCTLNWPIEAFPAPEEVNYTKKIKLSGLALDHKVTGDLFYTH VTTMGQRLSQKAPSLEDGSDAFMSPQDVRGTSENLPERSVPLRKSLCSPTFLWSLLTMGMTQLRIIF YMAAVNKMLEYLVTGGQEHETNEQQQKVAETVGFYSSVFGAMQLLCLLTCPLIGYIMDWRIKDCVDA PTQGTVLGDARDGVATKSIRPRYCKIQKLTNAISAFTLTNLLLVGFGITCLINNLHLQFVTFVLHTI VRGFFHSACGSLYAAVFPSNHFGTLTGLQSLISAVFALLQQPLFMAMVGPLKGEPFWVNLGLLLFSL LGFLLPSYLFYYRARLQQEYAANGMGPLKVLSGSEVTA		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 33B.

Table 33B. Comparison of NOV33a against NOV33b.		
Protein Sequence	**NOV33a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV33b	1..514 1..574	494/574 (86%) 494/574 (86%)

Further analysis of the NOV33a protein yielded the following properties shown in Table 33C.

Table 33C. Protein Sequence Properties NOV33a	
PSort analysis:	0.6450 probability located in mitochondrial inner membrane; 0.6000 probability located in plasma membrane; 0.5634 probability located in mitochondrial intermembrane space; 0.4367 probability located in mitochondrial matrix space
SignalP analysis:	Cleavage site between residues 45 and 46

A search of the NOV33a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 33D.

Table 33D. Geneseq Results for NOV33a				
Geneseq	**Protein/Organism/Length**	**NOV33a**	**Identities/**	**Expect**

Identifier	[Patent #, Date]	Residues/ Match Residues	Similarities for the Matched Region	Value
AAY44897	Human PB39 protein dysregulated in prostate cancer - Homo sapiens, 559 aa. [WO200005376-A1, 03-FEB-2000]	1..514 1..559	514/559 (91%) 514/559 (91%)	0.0
AAW64554	Human liver cell clone HP10301 protein - Homo sapiens, 559 aa. [WO9821328-A2, 22-MAY-1998]	1..514 1..559	514/559 (91%) 514/559 (91%)	0.0
AAY44898	Human PB39 variant protein dysregulated in prostate cancer - Homo sapiens, 560 aa. [WO200005376-A1, 03-FEB-2000]	1..467 1..512	467/512 (91%) 467/512 (91%)	0.0
AAB94537	Human protein sequence SEQ ID NO:15277 - Homo sapiens, 485 aa. [EP1074617-A2, 07-FEB-2001]	68..514 39..485	447/447 (100%) 447/447 (100%)	0.0
AAE05505	Mature human HC-like protein #2 - Homo sapiens, 529 aa. [WO200155435-A2, 02-AUG-2001]	68..495 85..505	250/436 (57%) 320/436 (73%)	e-138

In a BLAST search of public sequence datbases, the NOV33a protein was found to have homology to the proteins shown in the BLASTP data in Table 33E.

Table 33E. Public BLASTP Results for NOV33a				
Protein Accession Number	Protein/Organism/Length	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O75387	PB39 (Prostate cancer OVEREXPRESSED gene 1) - Homo sapiens (Human), 559 aa.	1..514 1..559	514/559 (91%) 514/559 (91%)	0.0
Q9D0H7	2610016F07Rik protein - Mus musculus (Mouse), 654 aa.	11..512 101..652	417/552 (75%) 453/552 (81%)	0.0
AAH27923	Hypothetical 62.7 kDa protein - Homo sapiens (Human), 569 aa.	1..495 1..545	297/560 (53%) 374/560 (66%)	e-154

BAC11450	CDNA FLJ90692 fis, clone PLACE1006443, weakly similar to Homo sapiens PB39 mRNA - Homo sapiens (Human), 460 aa.	68..495 16..436	250/436 (57%) 320/436 (73%)	e-138
BAC11383	CDNA FLJ90587 fis, clone PLACE1000914, weakly similar to Homo sapiens PB39 mRNA - Homo sapiens (Human), 460 aa.	68..495 16..436	249/436 (57%) 320/436 (73%)	e-137

PFam analysis predicts that the NOV33a protein contains the domains shown in the Table 33F.

Table 33F. Domain Analysis of NOV33a

Pfam Domain	NOV33a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 34.

The NOV34 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 34A.

Table 34A. NOV34 Sequence Analysis

	SEQ ID NO: 131	458 bp	
NOV34a, CG149777-01 DNA Sequence	AATCGCCTTACATGATGTGGCCCATGCACACCCCACTGCTGCTGCTGACTGCCTTGATGGTGGCCGT GGCCGGGAGTGCCTCGGCCCAATCTAGGACCTTGGCAGGTGGCATCCATGCCACAGACCTCAATGAC AAGAGTGTGCAGCGTGCCCTGGACTTTGCCATCAGCGAGTACAACAAGGTCATTAATAAGGATGAGT ACTACAGCCGCCCTCTGCAGGTGATGGCTGCCTACCAGCAGATCGTGGGTGGGGTGAACTACTACTT CAATGTGAAGTTCGGTCGAACCACATGCACCAAGTCCCAGCCCAACTTGGACAACTGTCCCTTCAAT GACCAGCCAAAACTGAAAGAGGAAGAGTTCTGCTCTTTCCAGATCAATGAAGTTCCCTGGGAGGATA AAATTTCCATTCTGAACTACAAGTGCCGGAAAGTCTAGGGGTCTGTGCAAGGCCTG		
	ORF Start: ATG at 12		ORF Stop: TAG at 438

	SEQ ID NO: 132	142 aa	MW at 16133.4kD
NOV34a, CG149777-01 Protein Sequence	MMWPMHTPLLLLTALMVAVAGSASAQSRTLAGGIHATDLNDKSVQRALDFAISEYNKVINKDEYYSR PLQVMAAYQQIVGGVNYYFNVKFGRTTCTKSQPNLDNCPFNDQPKLKEEEFCSFQINEVPWEDKISI LNYKCRKV		

	SEQ ID NO: 133	285 bp	
NOV34b,	AACATGATGTGGCCCATGCACACCCCACTGCTGCTGCTGACTGCCTTGATGGTGGCCGTGGCCGGGA GTGCCTCGGCCCAATCTAGGACCTTGGCAGGTGGCATCCATGCCACAGACCTCAATGACAAGAGTGT		

CG149777-02 DNA Sequence	GCAGCGTGCCCTGGACTTTGCCTTCAATGACCAGCCAAAACTGAAAGAGGAAGAGTTCTGCTCTTTC CAGATCAATGAAGTTCCCTGGGAGGATAAAATTTCCATTCTGAACTACAAGTGCCGGAAAGTCTAGG GGTCTGTGCAAGGCCTG		
	ORF Start: ATG at 4		ORF Stop: TAG at 265

	SEQ ID NO: 134	87 aa	MW at 9781.2kD
NOV34b, CG149777-02 Protein Sequence	MMWPMHTPLLLLTALMVAVAGSASAQSRTLAGGIHATDLNDKSVQRALDFAFNDQPKLKEEEFCSFQ INEVPWEDKISILNYKCRKV		

	SEQ ID NO: 135	280 bp
NOV34c, 257474374 DNA Sequence	CACCAAGCTTATGATGTGGCCCATGCACACCCCACTGCTGCTGCTGACTGCCTTGATGGTGGCCGTG GCCGGGAGTGCCTCGGCCCAATCTAGGACCTTGGCAGGTGGCATCCATGCCACAGACCTCAATGACA AGAGTGTGCAGCGTGCCCTGGACTTTGCCTTCAATGACCAGCCAAAACTGAAAGAGGAAGAGTTCTG CTCTTTCCAGATCAATGAAGTTCCCTGGGAGGATAAAATTTCCATTCTGAACTACAAGTGCCGGAAA GTCCTCGAGGGC	
	ORF Start: at 2	ORF Stop: end of sequence

	SEQ ID NO: 136	93 aa	MW at 10423.0kD
NOV34c, 257474374 Protein Sequence	TKLMMWPMHTPLLLLTALMVAVAGSASAQSRTLAGGIHATDLNDKSVQRALDFAFNDQPKLKEEEFC SFQINEVPWEDKISILNYKCRKVLEG		

	SEQ ID NO: 137	205 bp
NOV34d, 257474386 DNA Sequence	CACCAAGCTTCAATCTAGGACCTTGGCAGGTGGCATCCATGCCACAGACCTCAATGACAAGAGTGTG CAGCGTGCCCTGGACTTTGCCTTCAATGACCAGCCAAAACTGAAAGAGGAAGAGTTCTGCTCTTTCC AGATCAATGAAGTTCCCTGGGAGGATAAAATTTCCATTCTGAACTACAAGTGCCGGAAAGTCCTCGA GGGC	
	ORF Start: at 2	ORF Stop: end of sequence

	SEQ ID NO: 138	68 aa	MW at 7827.8kD
NOV34d, 257474386 Protein Sequence	TKLQSRTLAGGIHATDLNDKSVQRALDFAFNDQPKLKEEEFCSFQINEVPWEDKISILNYKCRKVLE G		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 34B.

Table 34B. Comparison of NOV34a against NOV34b through NOV34d.

Protein Sequence	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV34b	1..142 1..87	67/142 (47%) 69/142 (48%)

NOV34c	1..142 4..90	67/142 (47%) 69/142 (48%)
NOV34d	26..142 4..65	58/117 (49%) 60/117 (50%)

Further analysis of the NOV34a protein yielded the following properties shown in Table 34C.

Table 34C. Protein Sequence Properties NOV34a	
PSort analysis:	0.7857 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 26 and 27

A search of the NOV34a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 34D.

Table 34D. Geneseq Results for NOV34a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV34a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAO15149	Human cystatin D protein sequence - Homo sapiens, 142 aa. [US2002052476-A1, 02-MAY-2002]	1..142 1..142	142/142 (100%) 142/142 (100%)	2e-80
AAE02408	Human cystatin D precursor protein - Homo sapiens, 142 aa. [US6235708-B1, 22-MAY-2001]	1..142 1..142	142/142 (100%) 142/142 (100%)	2e-80
AAE04437	Human cystatin D homologue protein - Homo sapiens, 142 aa. [US6245529-B1, 12-JUN-2001]	1..142 1..142	142/142 (100%) 142/142 (100%)	2e-80
AAE11210	Human cystatin D (CysD) protein - Homo sapiens, 142 aa. [US6300477-B1, 09-OCT-2001]	1..142 1..142	142/142 (100%) 142/142 (100%)	2e-80

AAY81137	Human wild-type cystatin D - Homo sapiens, 122 aa. [WO200008159-A2, 17-FEB-2000]	21..142 1..122	122/122 (100%) 122/122 (100%)	3e-68

In a BLAST search of public sequence datbases, the NOV34a protein was found to have homology to the proteins shown in the BLASTP data in Table 34E.

Table 34E. Public BLASTP Results for NOV34a

Protein Accession Number	Protein/Organism/Length	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
A47142	cystatin D precursor - human, 142 aa.	1..142 1..142	142/142 (100%) 142/142 (100%)	7e-80
P28325	Cystatin D precursor - Homo sapiens (Human), 142 aa.	1..142 1..142	141/142 (99%) 141/142 (99%)	6e-79
P09228	Cystatin SA precursor (Cystatin S5) - Homo sapiens (Human), 141 aa.	1..141 1..140	80/141 (56%) 108/141 (75%)	2e-42
P01036	Cystatin S precursor (Salivary acidic protein-1) (Cystatin SA-III) - Homo sapiens (Human), 141 aa.	1..141 1..140	79/141 (56%) 109/141 (77%)	2e-41
P01037	Cystatin SN precursor (Salivary cystatin SA-1) (Cystain SA-I) - Homo sapiens (Human), 141 aa.	5..141 5..140	78/137 (56%) 105/137 (75%)	5e-40

PFam analysis predicts that the NOV34a protein contains the domains shown in the Table 34F.

Table 34F. Domain Analysis of NOV34a

Pfam Domain	NOV34a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cystatin	32..138	45/113 (40%) 99/113 (88%)	1.6e-39

Example 35.

The NOV35 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 35A.

Table 35A. NOV35 Sequence Analysis			
	SEQ ID NO: 139	1733 bp	
NOV35a, CG150005-01 DNA Sequence	GGACACCGTGCGTACCGGCCTGCGGCGCCCGGCCACCGGGGCGGACCGCGGAACCCGAGGCCATGTC GGACACCGTGCGTACCGGCCTGCGGCGCCCGGCCACCGGGGCGGACCGCGGAACCCGAGGCCATGTC CCATGAAAAGAGTTTTTGGTGTCTGGGGACAACTATCCTCCCCCCAACCCTGGATATCCGGGGGG CCCCAGCCACCCATGCCCCCCTATGCTCAGCCTCCCTACCCTGGGGCCCCTTACCCACAGCCCCCTT TCCAGCCCTCCCCCTACGGTCAGCCAGGGTACCCCCATGGCCCCAGCCCCTACCCCCAAGCCCTACC CCCAGGGCCCCTACCCACAAGAGGGCTACCCACAGGGCCCCTACCCCCAGAGCCCCTTCCCCCCCAA CCCCTATGGACAGCCATTCCCAGGACAAGACCCTGACTCACCCCAGCATGGAAACTACCAGGAGGAG GGTCCCCCATCCTACTATGACAACCAGGACTTCCCTGCCACCAACTGGGATAAGAGCATCCGACAGG CCTTCATCCGCAAGGTGTTCCTAGTGCTGACCTTGCAGCTGTCGGTGACCCTGTCCACGGTGTCTGT GTTCACTTTTGTTGCGGAGGTGAAGGGCTTTGTCCGGGAGAATGTCTGGACCTACTATGTCTCCTAT GCTGTCTTCTTCATCTCTCTCATCGTCCTCAGCTGTTGTGGGGACTTCCGGCGAAAGCACCCCTGGA ACCTTGTTGCACTGTCGGTCCTGACCGCCAGCCTGTCGTACATGGTGGGGATGATCGCCAGCTTCTA CAACACCGAGGCAGTCATCATGGCCGTGGGCATCACCACAGCCGTCTGCTTCACCGTCGTCATCTTC TCCATGCAGACCCGCTACGACTTCACCTCATGCATGGGCGTGCTCCTGGTGAGCATGGTGGTGCTCT TCATCTTCGCCATTCTCTGCATCTTCATCCGGAACCGCATCCTGGAGATCGTGTACGCCTCAACTGG GGCTCTGCTGCTGACCTGCTTCCTCGCAGTGGACACCCAGCTGCTGCTGGGGAACAAGCAGCTGTCC CTGAGCCCAGAAGAGTATGTGTTTGCTGCGCTGAACCTGTACACAGACATCATCAACATCTTCCTGT ACATCCTCACCATCATTGGCCCGCCAAGGAGTAGCCGAGCTCCAGCTCGCTGTGCCCGCTCAGGTGG CACGGCTGCCCCTGGCACGGCAGTGCCAGCTGTACTTCCCCTCTCTCTTGTCCCCAGGCACAGCCTA GGGAAAAGGATGCCTCTCTCCAACCCTCCTGTATGTACACTGCAGATACTTCCATTTGGACCCGCTG TGGCCACAGCATGGGCCCCTTTAGTCCTCCCGCCCCCGCCAAGGGGCACCAAGGCCACGTTTCCGTG CCACCTCCTGTCTACTCATTGTTGCATGAGCCCTGTCTGCCAGCTTCCACCCCAGGGACTGGGGGTC AGCGAACAGGTCCAAGGATTGAGCTCAATGGGTGAGGGTGCACGTCTTCCCTCCTGTCCCAGCTCCC CAGCCTGGCGTAGAGCACCCCTCCCCTCCCCCCCAAGTGCTGCCCTCTGGGGACATGGCGGAGTGGG GGTCTTATCCCCTGAGGGCAGAGGATGGCATGTTTCAGGGGAGAGAGGAAGCCTTCCTCTCAATTTG TTGTCAGTGAAATTCCAA**TAA**ATGGGATTTGCTCTCTGCAAAAAAAAAAAAAAAAAAAAAAAAAGGA AGCAAAGCCCCCAACCGACAGCACCATCAAATCAGCAACTGACAACCGACCGACACCA		
	ORF Start: ATG at 70		ORF Stop: TAA at 1627

	SEQ ID NO: 140	519 aa	MW at 56107.8kD
NOV35a, CG150005-01 Protein Sequence	MKRVFWCLGTTILPPTLDIRGGPSHPCPPMLSLPTLGPLTHSPLSSPPPTVSQGTPMAPAPTPKPYP QGPYPQEGYPQGPYPQSPFPPNPYGQPFPGQDPDSPQHGNYQEEGPPSYYDNQDFPATNWDKSIRQA FIRKVFLVLTLQLSVTLSTVSVFTFVAEVKGFVRENVWTYYVSYAVFFISLIVLSCCGDFRRKHPWN LVALSVLTASLSYMVGMIASFYNTEAVIMAVGITTAVCFTVVIFSMQTRYDFTSCMGVLLVSMVVLF IFAILCIFIRNRILEIVYASTGALLLTCFLAVDTQLLLGNKQLSLSPEEYVFAALNLYTDIINIFLY ILTIIGPPRSSRAPARCARSGGTAAPGTAVPAVLPLSLVPRHSLGKRMPLSNPPVCTLQILPFGPAV ATAWAPLVLPPPPRGTKATFPCHLLSTHCCMSPVCQLPPQGLGVSEQVQGLSSMGEGARLPSCPSSP AWRRAPLPSPPSAALWGHGGVGVLSPEGRGWHVSGERGSLPLNLLSVKFQ		

Further analysis of the NOV35a protein yielded the following properties shown in Table 35B.

Table 35B. Protein Sequence Properties NOV35a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.5510 probability located in mitochondrial inner membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 22 and 23

A search of the NOV35a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 35C.

Table 35C. Geneseq Results for NOV35a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW62612	Human glutamate-binding protein (HGLUBP) - Homo sapiens, 369 aa. [WO9821241-A1, 22-MAY-1998]	1..341 1..365	337/365 (92%) 337/365 (92%)	0.0
ABB12050	Human leukocyte HP00804 protein homologue, SEQ ID NO:2420 - Homo sapiens, 461 aa. [WO200157188-A2, 09-AUG-2001]	49..496 1..461	339/463 (73%) 350/463 (75%)	e-180
AAW64535	Human leukocyte cell clone HP00804 protein - Homo sapiens, 371 aa. [WO9821328-A2, 22-MAY-1998]	2..341 4..367	293/364 (80%) 297/364 (81%)	e-161
AAY48255	Human prostate cancer-associated protein 41 - Homo sapiens, 321 aa. [DE19811193-A1, 16-SEP-1999]	50..328 1..304	240/304 (78%) 246/304 (79%)	e-129
ABB60180	Drosophila melanogaster polypeptide SEQ ID NO 7332 - Drosophila melanogaster, 324 aa. [WO200171042-A2, 27-SEP-2001]	67..344 36..323	134/291 (46%) 190/291 (65%)	1e-65

In a BLAST search of public sequence datbases, the NOV35a protein was found to have homology to the proteins shown in the BLASTP data in Table 35D.

Table 35D. Public BLASTP Results for NOV35a

Protein Accession Number	Protein/Organism/Length	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

S19586	N-methyl-D-aspartate receptor glutamate-binding chain - rat, 516 aa.	1..513 1..516	380/517 (73%) 403/517 (77%)	0.0
Q63863	NMDA receptor glutamate-binding subunit - Rattus sp, 516 aa.	1..513 1..516	379/517 (73%) 402/517 (77%)	0.0
Q9ESF4	LAG protein - Mus musculus (Mouse), 345 aa.	22..341 21..341	277/322 (86%) 287/322 (89%)	e-158
O43836	NMDA receptor glutamate-binding chain - Homo sapiens (Human), 208 aa (fragment).	197..399 6..208	172/207 (83%) 178/207 (85%)	4e-83
AAM68613	CG3798-PA - Drosophila melanogaster (Fruit fly), 313 aa.	67..344 25..312	134/291 (46%) 190/291 (65%)	3e-65

PFam analysis predicts that the NOV35a protein contains the domains shown in the Table 35E.

Table 35E. Domain Analysis of NOV35a			
Pfam Domain	**NOV35a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
UPF0005	157..344	76/208 (37%) 180/208 (87%)	7.9e-79

Example 36.

The NOV36 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 36A.

Table 36A. NOV36 Sequence Analysis	
	SEQ ID NO: 141 1675 bp
NOV36a, CG150189-01 DNA Sequence	ATGGAGGGCGCAGGGCCCCGGGGGGCCGGGCCGGCGCGGCGCCGGGGAGCCGGGGGGCCGCCGTCAC CGCTGCTGCCGTCGCTGCTGCTGCTGCTGCTGCTCTGGATGCTGCCGGACACCGTGGCGCCTCAGGA ACTGAACCCTCGCGGCCGCAACGTGTGCCGTGCTCCCGGCTCCCAGGTGCCCACGTGCTGCGCTGGC TGGAGGCAGCAAGGGGACGAGTGTGGATTGCGGTGTGCGAAGGCAACTCCACGTGCTCAGAGAACG AGGTGTGCGTGAGGCCTGGCGAGTGCCGCTGCCGCCACGGCTACTTCGGTGCCAACTGCGACACCAA GTGCCCGCGCCAGTTCTGGGGCCCCGACTGCAAGGAGCTGTGTAGCTGCCACCCACACGGGCAGTGC GAGGACGTGACAGGCCGGTGCAAGGGCCAGCAGCCGTGCACGGTGGCCGAGGGCCGCTGCTTGACGT GCGAGCCCGGCTGGAACGGAACCAAGTGCGACCAGCCTTGCGCCACCGGTTTCTATGGCGAGGGCTG CAGCCACCGCTGTCCGCCATGCCGCGACGGGCATGCCTGTAACCATGTCACCGGCAAGTGTACGCGC TGCAACGCGGGCTGGATCGGCGACCGGTGCGAGACCAAGTGTAGCAATGGCACTTACGGCGAGGACT GCGCCTTCGTGTGCGCCGACTGCGGCAGCGGACACTGCGACTTCCAGTCGGGGCGCTGCCTGTGCAG CCCTGGCGTCCACGGGCCCCACTGTAACGTGACGTGCCCGCCCGGACTCCACGGCGCGGACTGTGCT CAGGCCTGCAGCTGCCACGAGGACTCGTGCGACCCGGTCACTGGTGCCTGCCACCTAGAAACCAACC AGCGCAAGGGCGTGATGGGCGCGGGCGCGCTGCTCGTCCTGCTCGTCTGCCTGCTGCTCTCGCTGCT TGGCTGCTGCTGCGCTTGCCGCGGCAAGGACCCTACGCGCCGGGAGCTTTCGCTTGGGAGGAAGAAG GCGCCGCACCGACTATGCGGGCGCTTCAGTCGCATCAGCATGAAGCTGCCCCGGATCCCGCTCCGGA

	GGCAGAAACTACCCAAAGTCGTAGTGGCCCACCACGACCTGGATAACACACTCAACTGCAGCTTCCT GGAGCCACCCTCAGGGCTGGAGCAGCCCTCACCATCCTGGTCCTCTCGGGCCTCCTTCTCCTCGTTT GACACCACTGATGAAGGCCCTGTGTACTGTGTACCCCATGAGGGTAAGTAAGGCCCTACCTGGGCAT CACTCCAGCCCAGTGAAATGTTCCCATGGAAAAGCTGTGTTCTGGGTGGACACAGGAGAAGGGCAG GCAGCATGGAGAGGAAGGCCTTGGCCATGCTGGTACCTGAGGGTTGCCCACAGAGCTGAGGCCATAG AGCTGGACTCTGCTGCTCAGTACCGGAGACAGGTGTGGGGAGATGGGTAGGCCACAGCCCAGGGTTG CTCCTGGGGGAAAGTAGGCAGAGAGAAGTTTCTGGGCTTAGGTAGGGGGTGGCAGAGGAGACAGGAG GAAGGGATCCACAGAGTATGGAGTTGGATCCACACACAGCCTTTGATCCACAGATAGCAGAAAGGA GCCTGATGGTCTGGGATTCTGCCCCTAGAATTCAGCGGCCGCTTTTTTTTTTTTTTTTTTTTTTTT		
	ORF Start: ATG at 1		ORF Stop: TAA at 1255

	SEQ ID NO: 142	418 aa	MW at 44706.5kD
NOV36a, CG150189-01 Protein Sequence	MEGAGPRGAGPARRRGAGGPPSPLLPSLLLLLLLWMLPDTVAPQELNPRGRNVCRAPGSQVPTCCAG WRQQGDECGIAVCEGNSTCSENEVCVRPGECRCRHGYFGANCDTKCPRQFWGPDCKELCSCHPHGQC EDVTGRCKGQQPCTVAEGRCLTCEPGWNGTKCDQPCATGFYGEGCSHRCPPCRDGHACNHVTGKCTR CNAGWIGDRCETKCSNGTYGEDCAFVCADCGSGHCDFQSGRCLCSPGVHGPHCNVTCPPGLHGADCA QACSCHEDSCDPVTGACHLETNQRKGVMGAGALLVLLVCLLLSLLGCCCACRGKDPTRRELSLGRKK APHRLCGRFSRISMKLPRIPLRRQKLPKVVVAHHDLDNTLNCSFLEPPSGLEQPSPSWSSRASFSSF DTTDEGPVYCVPHEGK		

Further analysis of the NOV36a protein yielded the following properties shown in Table 36B.

Table 36B. Protein Sequence Properties NOV36a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 44 and 45

A search of the NOV36a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 36C.

Table 36C. Geneseq Results for NOV36a				
Geneseq Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV36a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM47668	MOL8b protein sequence - Homo sapiens, 865 aa. [WO200181578-A2, 01-NOV-2001]	64..416 202..564	298/370 (80%) 308/370 (82%)	0.0
AAM47667	MOL8a protein sequence - Homo sapiens, 884 aa. [WO200181578-A2, 01-NOV-2001]	64..416 271..633	298/370 (80%) 308/370 (82%)	0.0
AAB60394	Human nurse cell receptor	64..416	298/370 (80%)	0.0

	B6TNC#10a, SEQ ID NO:24 - Homo sapiens, 866 aa. [JP2000308492-A, 07-NOV-2000]	202..564	308/370 (82%)	
AAB60393	Human nurse cell receptor B6TNC#10, SEQ ID NO:21 - Homo sapiens, 866 aa. [JP2000308492-A, 07-NOV-2000]	64..416 202..564	298/370 (80%) 308/370 (82%)	0.0
AAB60395	Human nurse cell receptor B6TNC#10b, SEQ ID NO:26 - Homo sapiens, 871 aa. [JP2000308492-A, 07-NOV-2000]	64..416 202..569	298/375 (79%) 308/375 (81%)	0.0

In a BLAST search of public sequence datbases, the NOV36a protein was found to have homology to the proteins shown in the BLASTP data in Table 36D.

Table 36D. Public BLASTP Results for NOV36a

Protein Accession Number	Protein/Organism/Length	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96GP6	Unknown (Protein for IMAGE:4125591) - Homo sapiens (Human), 598 aa (fragment).	113..416 5..296	282/304 (92%) 286/304 (93%)	0.0
CAD29035	Sequence 17 from Patent WO0214358 - Homo sapiens (Human), 254 aa.	1..205 1..175	153/205 (74%) 158/205 (76%)	2e-88
BAC02696	SREC-5 - Homo sapiens (Human), 744 aa.	28..414 6..455	172/462 (37%) 224/462 (48%)	7e-76
Q14162	Endothelial cells scavenger receptor precursor (Acetyl LDL receptor) - Homo sapiens (Human), 830 aa.	35..414 154..541	153/405 (37%) 195/405 (47%)	2e-75
BAC02694	SREC-3 - Homo sapiens (Human), 569 aa.	35..367 154..497	130/355 (36%) 168/355 (46%)	1e-64

PFam analysis predicts that the NOV36a protein contains the domains shown in the Table 36E.

Table 36E. Domain Analysis of NOV36a

Pfam Domain	NOV36a Match Region	Identities/ Similarities for the Matched Region	Expect Value
laminin_EGF	80..122	13/60 (22%) 29/60 (48%)	0.13
laminin_EGF	126..183	19/66 (29%) 41/66 (62%)	0.039
laminin_EGF	186..228	16/60 (27%) 28/60 (47%)	0.33
laminin_EGF	231..271	18/60 (30%) 31/60 (52%)	0.22

Example 37.

The NOV37 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 37A.

Table 37A. NOV37 Sequence Analysis			
	SEQ ID NO: 143	1025 bp	
NOV37a, CG150267-01 DNA Sequence	ATTGTGTCTGGCGGCGGCGCGCGAGCCCACCGGCGGCTGCGGCGGGGCGGGAAGCCATGGAGCCGCG GGCGCTCGTCACGGCGCTCAGCCTCGGCCTCAGCCTGTGCTCCCTGGGGCTGCTCGTCACGGCCATC TTCACCGACCACTGGTACGAGACCGACCCCCGGCGCCACAAGGAGAGCTGCGAGCGCAGCCGCGCGG GCGCCGACCCCCCGGACCAGAAGAACCGCCTGATGCCGCTGTCGCACCTGCCGCTGCGGGACTCGCC CCCGCTGGGGCGCCGGCTGCTCCCGGGCGGCCCGGGGCGCGCCGACCCCGAGTCCTGGCGCTCGCTC CTGGGGCTCGGCGGGCTGGACGCCGAGTGCGGCCGGCCCCTCTTCGCCACCTACTCGGGCCTCTGGA GGAAGTGCTACTTCCTGGGCATCGACCGGGACATCGACACCCTCATCCTGAAAGGTATTGCGCAGCG ATGCACGGCCATCAAGTACCACTTTTCTCAGCCCATCCGCTTGCGAAACATTCCTTTTAATTTAACC AAGACCATACAGCAAGATGAGTGGCACCTGCTTCGGATATTTTGCACCATTTCCCTCTGTACTTATG CCGCCAGTATCTCGTATGATTTGAACCGGCTCCCAAAGCTAATTTATAGCCTGCCTGCTGATGTGGA ACATGGTTACAGCTGGTCCATCTTTTGCGCCTGGTGCAGTTTAGGCTTTATTGTGGCAGCTGGAGGT CTCTGCATCGCTTATCCGTTTATTAGCCGGACCAAGATTGCACAGCTAAAGTCTGGCAGAGACTCCA CGGTATGACTGTCCTCACTGGGCCTGTCCACAGTGCGAGCGACTCCTGAGGGGAACAGCGCGGAGTT CAGGAGTCCAAGCACAAAGCGGTCTTTTACATTCCAACCTGTTGCCTGCCAGCCCTTTCTGGATTAC TGATAGAAAATCATGCAAAACCTCCCAACCTTTCTAAGGACAAGACTACTGTGGATTCAAGTGCTTT AATGACTATTTATGCGTTGA		
	ORF Start: ATG at 57		ORF Stop: TGA at 810

	SEQ ID NO: 144	251 aa	MW at 28111.1kD
NOV37a, CG150267-01 Protein Sequence	MEPRALVTALSLGLSLCSLGLLVTAIFTDHWYETDPRRHKESCERSRAGADPPDQKNRLMPLSHLPL RDSPPLGRRLLPGGPGRADPESWRSLLGLGGLDAECGRPLFATYSGLWRKCYFLGIDRDIDTLILKG IAQRCTAIKYHFSQPIRLRNIPFNLTKTIQQDEWHLLRIFCTISLCTYAASISYDLNRLPKLIYSLP ADVEHGYSWSIFCAWCSLGFIVAAGGLCIAYPFISRTKIAQLKSGRDSTV		

Further analysis of the NOV37a protein yielded the following properties shown in Table 37B.

Table 37B. Protein Sequence Properties NOV37a

PSort analysis:	0.4600 probability located in plasma membrane; 0.3000 probability located in lysosome (membrane); 0.2800 probability located in endoplasmic reticulum (membrane); 0.2196 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 26 and 27

A search of the NOV37a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 37C.

Table 37C. Geneseq Results for NOV37a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG61908	Prostate cancer-associated protein #109 - Mammalia, 297 aa. [WO200230268-A2, 18-APR-2002]	1..251 1..297	250/297 (84%) 250/297 (84%)	e-142
AAB88388	Human membrane or secretory protein clone PSEC0131 - Homo sapiens, 297 aa. [EP1067182-A2, 10-JAN-2001]	1..251 1..297	250/297 (84%) 250/297 (84%)	e-142
AAE21272	Human gene 16 encoded secreted protein fragment, SEQ ID NO:138 - Homo sapiens, 207 aa. [WO200216390-A1, 28-FEB-2002]	92..251 2..207	159/206 (77%) 159/206 (77%)	4e-85
ABG64865	Human albumin fusion protein #1540 - Homo sapiens, 115 aa. [WO200177137-A1, 18-OCT-2001]	173..251 37..115	79/79 (100%) 79/79 (100%)	4e-41
ABB90241	Human polypeptide SEQ ID NO 2617 - Homo sapiens, 115 aa. [WO200190304-A2, 29-NOV-2001]	173..251 37..115	79/79 (100%) 79/79 (100%)	4e-41

In a BLAST search of public sequence datbases, the NOV37a protein was found to have homology to the proteins shown in the BLASTP data in Table 37D.

Table 37D. Public BLASTP Results for NOV37a

Protein Accession Number	Protein/Organism/Length	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC39753	Sequence 143 from Patent EP1067182 - Homo sapiens (Human), 297 aa.	1..251 1..297	250/297 (84%) 250/297 (84%)	e-142
AAH29530	Similar to RIKEN cDNA 2810417M05 gene - Homo sapiens (Human), 297 aa.	1..251 1..297	249/297 (83%) 250/297 (83%)	e-142
Q9CZ16	2810417M05Rik protein - Mus musculus (Mouse), 241 aa.	1..194 1..240	188/240 (78%) 191/240 (79%)	e-104
BAC11344	CDNA FLJ90516 fis, clone NT2RP3004481, weakly similar to BUTYROPHILIN PRECURSOR - Homo sapiens (Human), 388 aa.	10..64 323..376	23/60 (38%) 31/60 (51%)	0.43
CAC35426	Sequence 1 from Patent WO0118204 - Homo sapiens (Human), 340 aa.	10..64 275..328	23/60 (38%) 31/60 (51%)	0.43

PFam analysis predicts that the NOV37a protein contains the domains shown in the Table 37E.

Table 37E. Domain Analysis of NOV37a

Pfam Domain	NOV37a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 38.

The NOV38 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 38A.

Table 38A. NOV38 Sequence Analysis

	SEQ ID NO: 145	6094 bp	
NOV38a, CG150362-01 DNA Sequence	CAGGTGGGCGGGCTGGTGGGCAGAAGGGCAGACGGGCAGAGGAAGTGCCAGTGCCACTGGGACCATG GCTCTGACGGTAAGCGTGCAACGACTAACAGGGCTGACCGGCACCCACGACCGACAAGTGAAGCTCA CCTTTCGAGGCTTTACCCAGAAAACAAGAAAAATTCACTGTGGTCCAGAAGCAGATATCGGTGAGCT GTTCCGATGGCCCCACTATGGGGCTCCACTGGCTGGGGAGTGTCTGTCTGTGCAGGTGGTCAACTGC AGCCGTGTATTCAGCCTTAGGCCTCTAGGGACCCTGGTGATCTCCCTGCAGCAGCTACAGAATGCTG GGCATTTGGTGCTACGGGAAGCCCTAGTGGATGAGAATCTTCAAGTGTCCCCGATCCAGGTGGAGCT TGACCTGAAGTACCAGCCCCCAGAGGGCGCTACTGGAGCCTGGTCAGAGGAGGACTTTGGGGCACCC		

	ATCCAGGACAGCTTCGAGTTAATCATCCCCAATGTGGGCTTCCAGGAACTGGAGCCTGGGGAGGCCC
	AGCTGGAGCGGCGGGCAGTGGCTCTAGGCCGCAGGCTAGCTCGAAGTCTAGGCCAGCAGGACGATGA
	AGAGAATGAGCTGGAGCTTGAGCTGGAGCAGGACCTGGATGATGAGCCTGACGTGGAACTTTCTGGT
	GTTATGTTCAGCCCCCTCAAGAGCCGCGCCAGGGCCCTGGCCCATGGGGATCCCTTCCAGGTGTCCA
	GAGCTCAAGACTTCCAGGTGGGAGTCACTGTGCTGGAAGCCCAGAAACTGGTGGGAGTCAACATTAA
	CCCCTATGTGGCCGTGCAAGTGGGGGGCAGCGCCGTGTGACCGCCACACAGCGTGGACCAGTTGC
	CCCTTCTACAATGAGTACTTCTTGTTCGAATTTCATGACACGCGGCTTCGTCTCCAAGACTTGCTGC
	TGGAGATCACGGTGAGTGGGGTAGGGGTGACCAGTGTCCTTCAGAGAAGGGGGGATGAGAAAGCTGC
	AGGACTAACACCACCTTCCCCCAAGGCTTTCCATTCGCAGACCCTCCCCTTTATGGCCACCCGGATA
	GGCACCTTCAGGATGGACCTGGGCATCATCTTGGACCAGCCAGATGGCCAGTTCTACCAAAGATGGG
	TTCCGCTGCATGATCCCCGAGACACCCGCGCCGGGACCAAGGGTTTCATTAAGGTCACCTTGTCCGT
	GAGGGCGCGCGGGGACCTGCCCCCTCCAATGCTACCCCCGGCCCCAGGGCACTGTTCGGACATCGAG
	AAGAACCTGCTCCTGCCGCGCGGGGTGCCCGCCGAGAGGCCATGGGCGCGGCTCCGCGTGCGCCTGT
	ACCGCGCCGAGGGGCTTCCCGCGCTGCGCCTGGGGCTGCTGGGCAGCCTGGTCCGCGCCCTGCACGA
	CCAGCGCGTCCTGGTGGAGCCCTATGTGCGGGTGTCTTTCCTGGGGCAGGAGGGCGAGACGTCGGTG
	AGCGCCGAGGCGGCGGCGCCCGAATGGAACGAGCAGCTGAGCTTCGTGGAGCTCTTCCCGCCGCTGA
	CGCGCAGCCTCCGCCTGCAGCTGCGGGACGACGCGCCCCTGGTCGACGCGGCACTCGCTACGCACGT
	GCCGGACCTGAGGCGGATCTCCCATCCGGGCCGCGCGGCGGGGTTTAACCCTACCTTCGGCCCGGCC
	TGGGTGCCCCTCTATGGCTCGCCCCCCGGCGCGGGGCTCCGGGATAGTCTTCAAGGTCTCAACGAAG
	GCGTTGGCCAAGGCATTTGGTTCCGCGGCCGCCTTCTGCTGGCTGTGTCCATGCAGGTGTTGGAAGG
	GAGAGCTGAACCTGAGCCTCCCCAGGCCCAGCAGGGGTCCACGTTGTCCCGGCTCACCCGAAAGAAG
	AAAAAGAAAGCCAGAAGGGATCAGACCCCAAAGGCGGTTCCGCAGCACTTGGACGCCAGCCCCGGTG
	CCGAGGGGCCTGAGATCCCCCGTGCCATGGAGGTGGAGGTGGAGGAGCTGCTGCCGCTGCCAGAGAA
	TGTCCTGGCGCCCTGTGAAGATTTCCTGCTTTTCGGTGTGCTCTTCGAGGCCACCATGATCGACCCC
	ACCGTGGCCTCCCAGCCCATCAGCTTCGAGATCTCCATTGGTCGCGCAGGCCGTCTGGAGGAGCAAT
	TGGGCCGAGGGTCCAGGGCTGGGGAGGGAACTGAGGGTGCAGCCGTGGAGGCTCAGCCTCTGCTGGG
	AGCCAGGCCAGAGGAGGAGAAAGAGGAGGAAGAACTGGGGACCCATGCTCAGCGGCCTGAGCCCATG
	GACGGCAGTGGGCCATACTTCTGCTTGCCCCTCTGTCACTGCAAGCCATGCATGCATGTGTGGAGTT
	GCTGGGAGGACCACACCTGGCGCCTGCAGAGCAGCAACTGCGTGCGCAAAGTGGCCGAGAGGCTGGA
	CCAGGGGCTGCAGGAGGTTGAGAGACTGCAGCGCAAGCCGGGGCCTGGCGCCTGTGCACAGCTCAAG
	CAGGCACTGGAAGTACTGGTGGCTGGGAGCAGACAGTTTTGCCACGGTGCCGAGCGCAGGACGATGA
	CCCGGCCCAATGCCCTGGATCGATGCCGAGGGAAACTCCTGGTGCACAGCCTGAACCTTTTGGCTAA
	GCAAGGACTGCGACTTCTACGCAGCCTGAGACGGCGCAATGTGCAAAAGAAGGTGGCACTGGCCAAG
	AAGCTCCTGGCAAAACTGCGCTTTCTGGCTGAGGAGCCCCAGCCACCCCTCCCCGATGTGCTGGTCT
	GGATGCTCAGCGGGCAGCGCCGTGTGGCCTGGGCCCGGATCCCTGCCCAGGATGTGCTGTTCTCTGT
	GGTTGAGGAGGAACGGGGCCGAGACTGTGGAAGATCCAGAGTCTAATGCTCACGGCACCCGGGGCA
	GCCCCTGGTGAGGTCTGTGCCAAGCTGGAGCTCTTCCTGCGGCTGGGCCTGGGCAAGCAAGCCAAGG
	CCTGCACCTCTGAGCTGCCCCCGGATTTGCTGCCCGAGCCCTCAGCCGGGCTGCCCTCCAGCCTACA
	CCGGGACGACTTTAGCTACTTCCAACTCCGGGCTCACTTGTACCAGGCCCGGGGTGTGTTGGCTGCA
	GATGACAGTGGCCTCTCGGACCCCTTTGCTCGAGTCCTCATCTCTACCCAGTGTCAGACCACACGGG
	TCCTGGAGCAGACGCTGAGCCCTCTGTGGGATGAACTCCTGGTATTTGAGCAGTTGATCGTGGATGG
	GAGGAGGGAGCACCTGCAGGAGGAGCCTCCATTAGTGATCATCAATGTATTTGACCACAATAAGTTT
	GGCCCCCCCGTGTTCCTGGGCAGGGCACTGGCCGCCCCAAGGGTAAAGCTGATGGAGGACCCATACC
	AACGCCCAGAGTTGCAGTTCTTCCCCCTGAGGAAGGGACCCTGGGCAGCCGGAGAGCTCATTGCCGC
	CTTTCAACTCATTGAACTAGACTACAGTGGCCGACTTGAGCCCTCAGTGCCCAGTGAGGTGGAGCCC
	CAGGATCTGGCACCCCTGGTTGAGCCCCACTCTGGACGCCTGTCCCTTCCACCCAACGTGTGCCCAG
	TGCTCAGGGAGTTCCGTGTTGAGGTGCTGTTCTGGGGTCTTAGGGGACTTGGTCGTGTGCATCTGCT
	CGAGGTGGAGCAGCCCCAGGTTGTACTGGAGGTGGCTGGGCAAGGTGTGGAGTCTGAGGTCCTGGCC
	AGCTACCGTGAGAGCCCCAATTTCACTGAGCTTGTCAGGCATCTGACAGTGGTCTTCAAAGACACAG
	CTCCTCTCTTCCACCCCCAGGACTTGCCGGAGCAGCCTTACTTGCAGCCTCCACTCAGCATCTTGGT
	GATTGAGCGCCGGGCCTTTGGCCACACAGTCCTTGTGGGTTCCCACATTGTCCCCCACATGCTGCGA
	TTCACATTTCGGGGTCATGAGGATCCTCCTGAGGAGGAAGGAGAGATGGAGGAGACAGGGGATATGA
	TGCCCAAGGGACCTCAAGGACAGAAGTCCCTGGATCCCTTCTTGGCTGAAGCGGGTATATCCAGACA
	GCTCCTGAAGCCTCCTCTGAAGAAGCTCCCACTAGGAGGCCTCCTAAATCAAGGCCCTGGGCTGGAG
	GAAGACATCCCAGATCCAGAGGAGCTCGACTGGGGGTCCAAGTACTATGCGTCGCTGCAGGAGCTCC
	AGGGGCAGCACAACTTTGATGAAGATGAAATGGATGATCCTGGAGATTCAGATGGGGTCAACCTCAT
	TTCTATGGTTGGGGAGATCCAAGACCAGGGTGAGGCTGAAGTCAAAGGCACTGTGTCCCCAAAAAAA
	GCAGTTGCCACCCTGAAGATCTACAACAGGTCCCTGAAGGAAGAATTTAACCACTTTGAAGACTGGC
	TGAATGTGTTTCCTCTGTACCGAGGGCAAGGGGGCCAGGATGGAGGTGGAGAAGAGGAAGGATCTGG
	ACACCTTGTGGGCAAGTTCAAGGGCTCCTTCCTCATTTACCCTGAATCAGAGGCAGTGTTGTTCTCT
	GAGCCCCAGATCTCCCGGGGGATCCCACAGAACCGGCCCATCAAGCTCCTGGTCAGAGTGTATGTTG
	TAAAGGCTACCAACCTGGCTCCTGCAGACCCCAATGGCAAAGCAGACCCTTACGTGGTGGTGAGCGC
	TGGCCGGGAGCGGCAGGACACCAAGGAACGCTACATCCCCAAGCAGCTCAACCCCATCTTTGGAGAG
	ATCCTGGAGCTAAGCATCTCTCTCCCAGCTGAGACGGAGCTGACGGTCGCCGTATTTGATCATGACC
	TCGTGGGTTCTGACGACCTCATCGGGGAGACCCACATTGATCTGGAAAACCGATTCTATAGCCACCA
	CAGAGCAAACTGTGGGCTGGCCTCCCAGTATGAAGTAGATGGTTACAATGCCTGGCGTGATGCATTC
	TGGCCTTCGCAGATCCTGGCGGGGCTGTGCCAACGCTGTGGCCTCCCTGCCCCTGAATACCGAGCCG
	GTGCTGTCAAGGTGGGCAGCAAAGTCTTCCTGACACCACCGGAGACCCTGCCCCCAGTGGCGAGCGG
	GGACCCTGAAGAGGCCCAGGCATTGCTTGTGCTGCGGCGCTGGCAGGAAATGCCGGGTTTTGGGATC
	CAGCTGGTACCCGAGCATGTAGAAACCAGGCCTCTCTACCATCCCCACAGCCCAGGGCTGCTACAGG
	GATCTCTTCACATGTGGATTGACATCTTTCCTCAAGATGTGCCTGCTCCACCCCCAGTTGACATCAA
	GCCTCGGCAGCCAATCAGCTATGAGCTCAGAGTTGTCATCTGGAACACGGAGGATGTGGTTCTGGAT
	GACGAGAATCCACTCACCGGAGAGATGTCGAGTGACATCTATGTGAAGAGCTGGGTGAAGGGGTTGG
	AGCATGACAAGCAGGAGACAGACGTTCACTTCAACTCCCTGACTGGGGAGGGGAACTTCAATTGGCG
	CTTTGTGTTCCGCTTTGACTACCTGCCCACGGAGCGGGAGGTGAGCGTCTGGCGCAGGTCTGGACCC
	TTTGCCCTGGAGGAGGCGGAGTTCCGGCAGCCTGCAGTGCTGGTCCTGCAGGTCTGGGACTATGACC

	GCATCTCTGCCAATGACTTCCTTGGATCCCTGGAGTTGCAGCTACCAGACATGGTGCGTGGGGCCCG GGGCCCCGAGCTCTGCTCTGTGCAGCTGGCCCGCAATGGGGCCGGGCCGAGGTGCAATCTGTTTCGC TGCCGCCGCCTGAGGGGCTGGTGGCCGGTAGTGAAGCTGAAGGAGGCAGAGGACGGCAAGGTGGAGG CAGAGTTTGAGCTGCTGACTGTGGAGGAGGCCGAGAAACGGCCAGTGGGGAAGGGGCGGAAGCAGCC AGAGCCTCTGGAGAAACCCAGCCGCCCCAAAACTTCCTTCAACTGGTTTGTGAACCCGCTGAAGACC TTTGTCTTCTTCATCTGGCGCCGGTACTGGCGCACCCTGGTGCTGCTGCTACTGGTGCTGCTCACCG TCTTCCTCCTCCTGGTCTTCTACACCATCCCTGGCCAGATCAGCCAGGTCATCTTCCGTCCCCTCCA CAAGTGACTCTCGCTGACCTTGGACACTCACCCAGGGTGCCAACCCTTCAATGCCTGCTCCTGG		
	ORF Start: ATG at 65		ORF Stop: TGA at 6035

	SEQ ID NO: 146	1990 aa	MW at 222395.9kD
NOV38a, CG150362-01 Protein Sequence	MALTVSVQRLTGLTGTHDRQVKLTFRGFTQKTRKIHCGPEADIGELFRWPHYGAPLAGECLSVQVVN CSRVFSLRPLGTLVISLQQLQNAGHLVLREALVDENLQVSPIQVELDLKYQPPEGATGAWSEEDFGA PIQDSFELIIPNVGFQELEPGEAQLERRAVALGRRLARSLGQQDDEENELELELEQDLDDEPDVELS GVMFSPLKSRARALAHGDPFQVSRAQDFQVGVTVLEAQKLVGVNINPYVAVQVGGQRRVTATQRGTS CPFYNEYFLFEFHDTRLRLQDLLLEITVSGVGVTSVLQRRGDEKAAGLTPPSPKAFHSQTLPFMATR IGTFRMDLGIILDQPDGQFYQRWVPLHDPRDTRAGTKGFIKVTLSVRARGDLPPPMLPPAPGHCSDI EKNLLLPRGVPAERPWARLRVRLYRAEGLPALRLGLLGSLVRALHDQRVLVEPYVRVSFLGQEGETS VSAEAAAPEWNEQLSFVELFPPLTRSLRLQLRDDAPLVDAALATHVPDLRRISHPGRAAGFNPTFGP AWVPLYGSPPGAGLRDSLQGLNEGVGQGIWFRGRLLLAVSMQVLEGRAEPEPPQAQQGSTLSRLTRK KKKKARRDQTPKAVPQHLDASPGAEGPEIPRAMEVEVEELLPLPENVLAPCEDFLLFGVLFEATMID PTVASQPISFEISIGRAGRLEEQLGRGSRAGEGTEGAAVEAQPLLGARPEEEKEEEELGTHAQRPEP MDGSGPYFCLPLCHCKPCMHVWSCWEDHTWRLQSSNCVRKVAERLDQGLQEVERLQRKPGPGACAQL KQALEVLVAGSRQFCHGAERRTMTRPNALDRCRGKLLVHSLNLLAKQGLRLLRSLRRRNVQKKVALA KKLLAKLRFLAEEPQPPLPDVLVWMLSGQRRVAWARIPAQDVLFSVVEEERGRDCGKIQSLMLTAPG AAPGEVCAKLELFLRLGLGKQAKACTSELPPDLLPEPSAGLPSSLHRDDFSYFQLRAHLYQARGVLA ADDSGLSDPFARVLISTQCQTTRVLEQTLSPLWDELLVFEQLIVDGRREHLQEEPPLVIINVFDHNK FGPPVFLGRALAAPRVKLMEDPYQRPELQFFPLRKGPWAAGELIAAFQLIELDYSGRLEPSVPSEVE PQDLAPLVEPHSGRLSLPPNVCPVLREFRVEVLFWGLRGLGRVHLLEVEQPQVVLEVAGQGVESEVL ASYRESPNFTELVRHLTVVFKDTAPLFHPQDLPEQPYLQPPLSILVIERRAFGHTVLVGSHIVPHML RFTFRGHEDPPEEEGEMEETGDMMPKGPQGQKSLDPFLAEAGISRQLLKPPLKKLPLGGLLNQGPGL EEDIPDPEELDWGSKYYASLQELQGQHNFDEDEMDDPGDSDGVNLISMVGEIQDQGEAEVKGTVSPK KAVATLKIYNRSLKEEFNHFEDWLNVFPLYRGQGGQDGGGEEEGSGHLVGKFKGSFLIYPESEAVLF SEPQISRGIPQNRPIKLLVRVYVVKATNLAPADPNGKADPYVVVSAGRERQDTKERYIPKQLNPIFG EILELSISLPAETELTVAVFDHDLVGSDDLIGETHIDLENRFYSHHRANCGLASQYEVDGYNAWRDA FWPSQILAGLCQRCGLPAPEYRAGAVKVGSKVFLTPPETLPPVASGDPEEAQALLVLRRWQEMPGFG IQLVPEHVETRPLYHPHSPGLLQGSLHMWIDIFPQDVPAPPPVDIKPRQPISYELRVVIWNTEDVVL DDENPLTGEMSSDIYVKSWVKGLEHDKQETDVHFNSLTGEGNFNWRFVFRFDYLPTEREVSVWRRSG PFALEEAEFRQPAVLVLQVWDYDRISANDFLGSLELQLPDMVRGARGPELCSVQLARNGAGPRCNLF RCRRLRGWWPVVKLKEAEDGKVEAEFELLTVEEAEKRPVGKGRKQPEPLEKPSRPKTSFNWFVNPLK TFVFFIWRRYWRTLVLLLLVLLTVFLLLVFYTIPGQISQVIFRPLHK		

Further analysis of the NOV38a protein yielded the following properties shown in Table 38B.

Table 38B. Protein Sequence Properties NOV38a	
PSort analysis:	0.8000 probability located in mitochondrial inner membrane; 0.7000 probability located in plasma membrane; 0.3793 probability located in microbody (peroxisome); 0.3500 probability located in nucleus
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV38a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 38C.

Table 38C. Geneseq Results for NOV38a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU70673	Human otoferlin #2 - Homo sapiens, 1997 aa. [WO200170972-A2, 27-SEP-2001]	1..1984 1..1994	811/2104 (38%) 1198/2104 (56%)	0.0
AAU70669	Murine cochlea otoferlin - Mus sp, 2298 aa. [WO200170972-A2, 27-SEP-2001]	1..1978 35..2017	797/2091 (38%) 1188/2091 (56%)	0.0
AAU70674	Murine otoferlin #2 - Mus sp, 1992 aa. [WO200170972-A2, 27-SEP-2001]	1..1978 1..1983	797/2091 (38%) 1188/2091 (56%)	0.0
AAU70675	Human otoferlin - Homo sapiens, 1230 aa. [WO200170972-A2, 27-SEP-2001]	803..1984 12..1227	540/1249 (43%) 769/1249 (61%)	0.0
AAU70672	Human otoferlin #1 - Homo sapiens, 1307 aa. [WO200170972-A2, 27-SEP-2001]	783..1984 49..1304	542/1289 (42%) 778/1289 (60%)	0.0

In a BLAST search of public sequence datbases, the NOV38a protein was found to have homology to the proteins shown in the BLASTP data in Table 38D.

Table 38D. Public BLASTP Results for NOV38a				
Protein Accession Number	**Protein/Organism/Length**	**NOV38a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9HC10	Otoferlin (Fer-1 like protein 2) - Homo sapiens (Human), 1997 aa.	1..1984 1..1994	810/2104 (38%) 1197/2104 (56%)	0.0
Q9ESF1	Otoferlin (Fer-1 like protein 2) - Mus musculus (Mouse), 1997 aa.	1..1984 1..1994	803/2100 (38%) 1188/2100 (56%)	0.0
Q9H4S7	BA563A22B.1 (Contains a novel protein similar to otoferlin (A FER-1-like protein)) - Homo sapiens (Human), 615 aa (fragment).	885..1499 1..615	615/615 (100%) 615/615 (100%)	0.0

Q9NTZ8	DJ309K20.1.1 (Novel protein similar to dysferlin, isoform 1) - Homo sapiens (Human), 552 aa (fragment).	231..782 1..552	552/552 (100%) 552/552 (100%)	0.0
Q9H448	DJ477O4.1.1 (Novel protein similar to otoferlin and dysferlin, isoform 1) - Homo sapiens (Human), 531 aa (fragment).	1500..1990 1..531	491/531 (92%) 491/531 (92%)	0.0

PFam analysis predicts that the NOV38a protein contains the domains shown in the Table 38E.

Table 38E. Domain Analysis of NOV38a

Pfam Domain	NOV38a Match Region	Identities/ Similarities for the Matched Region	Expect Value
C2	231..303	23/97 (24%) 49/97 (51%)	0.037
C2	421..515	26/107 (24%) 60/107 (56%)	0.0038
C2	993..1083	26/101 (26%) 61/101 (60%)	0.0015
C2	1493..1576	32/97 (33%) 57/97 (59%)	1.8e-11

Example 39.

The NOV39 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 39A.

Table 39A. NOV39 Sequence Analysis

	SEQ ID NO: 147	1293 bp	
NOV39a, CG150637-01 DNA Sequence	ATCAAATACCAGCTCTGCCAGTAAGAAGTTGCATCTCCCAGTGAAATGCTGCTGCTGCCATTTCAAC TGTTAGCTGTTCTCTTTCCTGGTGGTAACAGTGAACATGCCTTCCAGGGGCCGACCTCCTTTCATGT TATCCAGACCTCGTCCTTTACCAATAGTACCTGGGCACAAACTCAAGGCTCAGGCTGGTTGGATGAT TTGCAGATTCATGGCTGGGATAGCGACTCAGGCACTGCCATATTCCTGAAGCCTTGGTCTAAAGGTA ACTTTAGTGATAAGGAGGTTGCTGAGTTAGAGGAGATATTCCGAGTCTACATCTTTGGATTCGCTCG AGAAGTACAAGACTTTGCCGGTGATTTCCAGATGAAATACCCCTTTGAGATCCAGGGCATAGCAGGC TGTGAGCTACATTCTGGAGGTGCCATAGTAAGCTTCCTGAGGGGAGCTCTAGGAGGATTGGATTTCC TGAGTGTCAAGAATGCTTCATGTGTGCCTTCCCCAGAAGGTGGCAGCAGGGCACAGAAATTCTGTGC ACTAATCATACAATATCAAGGTATCATGGAAACTGTGAGAATTCTCCTCTATGAAACCTGCCCCCGA TATCTCTTGGGCGTCCTCAATGCAGGAAAAGCAGATCTGCAAAGACAAGTGAAGCCTGAGGCCTGGC TGTCCAGTGGCCCCAGTCCTGGACCTGGCCGTCTGCAGCTTGTGTGCCATGTCTCAGGATTCTACCC AAAGCCCGTGTGGGTGATGTGGATGCGGGGTGAGCAGGAGCAGCAGGGCACTCAGCTAGGGGACATC CTGCCCAATGCTAACTGGACATGGTATCTCCGAGCAACCCTGGATGTGGCAGATGGGGAGGCGGCTG GCCTGTCCTGTCGGGTGAAGCACAGCAGTTTAGAGGGCCAGGACATCATCCTCTACTGGAGAAACCC CACCTCCATTGGCTCAATTGTTTTGGCAATAATAGTGCCTTCCTTGCTCCTTTTGCTATGCCTTGCA TTATGGTATATGAGGCGCCGGTCATATCAGAATATCCCATGAGCCATCATCATGTCTCCTCTCCCAT		

	TCGCAATAAGCTACCAAGAAGCCCAAGATATCAGCCCAAAAATCAATCTTATCATATTTCAAATGAT TTTCAAATTTGATGAAATCAGAGTTTTCATGTATTTTAAAATAAATTATTATTTAAACATCAGCAAA AAAGTACTTAAAACTGTAAATTTATTATGAGACTGTACTAACAGTGTGATTCACCCTGATTTTACAC ACATTAAAATGTTAGAAAAA		
	ORF Start: ATG at 46		ORF Stop: TGA at 1045

	SEQ ID NO: 148	333 aa	MW at 36939.0kD
NOV39a, CG150637-01 Protein Sequence	MLLLPFQLLAVLFPGGNSEHAFQGPTSFHVIQTSSFTNSTWAQTQGSGWLDDLQIHGWDSDSGTAIF LKPWSKGNFSDKEVAELEEIFRVYIFGFAREVQDFAGDFQMKYPFEIQGIAGCELHSGGAIVSFLRG ALGGLDFLSVKNASCVPSPEGGSRAQKFCALIIQYQGIMETVRILLYETCPRYLLGVLNAGKADLQR QVKPEAWLSSGPSPGPGRLQLVCHVSGFYPKPVWVMWMRGEQEQQGTQLGDILPNANWTWYLRATLD VADGEAAGLSCRVKHSSLEGQDIILYWRNPTSIGSIVLAIIVPSLLLLLCLALWYMRRRSYQNIP		

	SEQ ID NO: 149	880 bp	
NOV39b, CG150637-02 DNA Sequence	CCCTTATGCTGCTGCTGCCATTTCAACTGTTAGCTGTTCTCTTTCCTGGTGGTAACAGTGAACATGC CTTCCAGGGGCCGACCTCCTTTCATGTTATCCAGACCTCGTCCTTTACCAATAGTACCTGGGCACAA ACTCAAGGCTCAGGCTGGTTGGATGATTTGCAGATTCATGGCTGGGATAGCGACTCAGGCACTGCCA TATTCCTGAAGCCTTGGTCTAAAGGTAACTTTAGTGATAAGGAGGTTGCTGAGTTAGAGGAGATATT CCGAGTCTACATCTTTGGATTCGCTCGAGAAGTACAAGACTTTGCCGGTGATTTCCAGATGAAATAC CCCTTTGAGATCCAGGGCATAGCAGGCTGTGAGCTACATTCTGGAGGTGCCATAGTAAGCTTCCTGA GGGGAGCTCTAGGAGGATTGGATTTCCTGAGTGTCAAGAATGCTTCATGTGTGCCTTCCCCAGAAGG TGGCAGCAGGGCACAGAAATTCTGTGCACTAATCATACAATATCAAGGTATCATGGAAACTGTGAGA ATTCTCCTCTATGAAACCTGCCCCCGATATCTCTTGGGCGTCCTCAATGCAGGAAAAGCAGATCTGC AAAGACAAGTGAAGCCTGAGGCCTGGCTGTCCAGTGGCCCCAGTCCTGGACCTGGCCGTCTGCAGCT TGTGTGCCATGTCTCAGGATTCTACCCAAAGCCCGTGTGGGTGATGTGGATGCGGGGAAACCCCACC TCCATTGGCTCAATTGTTTTGGCAATAATAGTGCCTTCCTTGCTCCTTTTGCTATGCCTTGCATTAT GGTATATGAGGCGCCGGTCATATCAGAATATCCCATGAGCCATCATCATGTCTCCTCTCCCATTCGC AATAAGTAC		
	ORF Start: ATG at 6		ORF Stop: TGA at 840

	SEQ ID NO: 150	278 aa	MW at 30739.2kD
NOV39b, CG150637-02 Protein Sequence	MLLLPFQLLAVLFPGGNSEHAFQGPTSFHVIQTSSFTNSTWAQTQGSGWLDDLQIHGWDSDSGTAIF LKPWSKGNFSDKEVAELEEIFRVYIFGFAREVQDFAGDFQMKYPFEIQGIAGCELHSGGAIVSFLRG ALGGLDFLSVKNASCVPSPEGGSRAQKFCALIIQYQGIMETVRILLYETCPRYLLGVLNAGKADLQR QVKPEAWLSSGPSPGPGRLQLVCHVSGFYPKPVWVMWMRGNPTSIGSIVLAIIVPSLLLLLCLALWY MRRRSYQNIP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 39B.

Table 39B. Comparison of NOV39a against NOV39b.		
Protein Sequence	**NOV39a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV39b	15..247 15..247	228/233 (97%) 228/233 (97%)

Further analysis of the NOV39a protein yielded the following properties shown in Table 39C.

Table 39C. Protein Sequence Properties NOV39a

PSort analysis:	0.4600 probability located in plasma membrane; 0.3000 probability located in lysosome (membrane); 0.2800 probability located in endoplasmic reticulum (membrane); 0.2404 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 19 and 20

A search of the NOV39a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 39D.

Table 39D. Geneseq Results for NOV39a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG13799	Novel human diagnostic protein #13790 - Homo sapiens, 681 aa. [WO200175067-A2, 11-OCT-2001]	35..196 518..680	93/163 (57%) 118/163 (72%)	5e-49
AAY94506	Chicken BFIV21 class I MHC protein - Gallus gallus, 355 aa. [US6075125-A, 13-JUN-2000]	114..329 114..326	61/221 (27%) 107/221 (47%)	4e-17
AAY94508	Chicken BFIV19v1 class I MHC protein - Gallus gallus, 338 aa. [US6075125-A, 13-JUN-2000]	114..329 97..309	60/221 (27%) 105/221 (47%)	1e-15
AAG00593	Human secreted protein, SEQ ID NO: 4674 - Homo sapiens, 64 aa. [EP1033401-A2, 06-SEP-2000]	1..64 1..64	39/64 (60%) 46/64 (70%)	1e-15
ABB08372	B-FIV*12 amino acid sequence - Gallus domesticus, 334 aa. [WO200194615-A2, 13-DEC-2001]	114..315 93..295	58/211 (27%) 99/211 (46%)	2e-15

In a BLAST search of public sequence datbases, the NOV39a protein was found to have homology to the proteins shown in the BLASTP data in Table 39E.

Table 39E. Public BLASTP Results for NOV39a

Protein Accession Number	Protein/Organism/Length	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P29016	T-cell surface glycoprotein CD1b precursor (CD1b antigen) - Homo sapiens (Human), 333 aa.	1..333 1..333	333/333 (100%) 333/333 (100%)	0.0
Q28565	T-cell surface glycoprotein CD1b-1 precursor (CD1b-1 antigen) (SCD1A25) - Ovis aries (Sheep), 333 aa.	1..332 1..332	248/332 (74%) 280/332 (83%)	e-150
Q29422	T-cell surface glycoprotein CD1b-2 precursor (CD1b-2 antigen) (SCD1B-42) (Antigen IAH-CC14) - Ovis aries (Sheep), 333 aa.	1..332 1..332	244/332 (73%) 282/332 (84%)	e-147
Q9GKE4	CD1B - Oryctolagus cuniculus (Rabbit), 332 aa.	1..332 1..331	237/332 (71%) 271/332 (81%)	e-140
Q9QZZ1	T-cell surface glycoprotein CD1b2 precursor (CD1-b2 antigen) - Cavia porcellus (Guinea pig), 332 aa.	1..332 1..331	228/332 (68%) 267/332 (79%)	e-134

PFam analysis predicts that the NOV39a protein contains the domains shown in the Table 39F.

Table 39F. Domain Analysis of NOV39a			
Pfam Domain	**NOV39a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ig	217..281	15/67 (22%) 45/67 (67%)	0.00019

Example 40.

The NOV40 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 40A.

Table 40A. NOV40 Sequence Analysis			
	SEQ ID NO: 151	505 bp	
NOV40a, CG150694-01	AATATGTCGCTCTTGGGACCCAAGGTGCTGCTGTTTCTTGCTGCATTCATCATCACCTCTGACTGGA TACCCCTGGGGGTCAATAGTCAACGAGGAGACGATGTGACTCAAGCGACTCCAGAAACATTCACAGA AGATCCTAATCTGGTGAATGATCCCGCTACAGATGAAACAGAGTGCTGGGATGAGAAATTTACCTGC		

DNA Sequence	ACAAGGCTCTACTCTGTGCATCGGCCGGTTAAACAATGCATTCATCAGTTATGCTTCACCAGTTTAC GACGTATGTACATCGTCAACAAGGAGATCTGCTCTCGTCTTGTCTGTAAGGAACACGAAGCTATGAA AGATGAGCTTTGCCGTCAGATGGCTGGTCTGCCCCCTAGGAGACTCCGTCGCTCCAATTACTTCCGA CTTCCTCCCTGTGAAAATGTGGATTTGCAGAGACCCAATGGTCTGTGATCATTGAAAAAGAGGAAAG AAGAAAAAAATGTATGGGTGAGAGGAAGGAGGATCTC		
	ORF Start: ATG at 4		ORF Stop: TGA at 448

	SEQ ID NO: 152	148 aa	MW at 17113.5kD
NOV40a, CG150694-01 Protein Sequence	MSLLGPKVLLFLAAFIITSDWIPLGVNSQRGDDVTQATPETFTEDPNLVNDPATDETECWDEKFTCT RLYSVHRPVKQCIHQLCFTSLRRMYIVNKEICSRLVCKEHEAMKDELCRQMAGLPPRRLRRSNYFRL PPCENVDLQRPNGL		

Further analysis of the NOV40a protein yielded the following properties shown in Table 40B.

Table 40B. Protein Sequence Properties NOV40a	
PSort analysis:	0.6850 probability located in plasma membrane; 0.6400 probability located in endoplasmic reticulum (membrane); 0.3700 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 29 and 30

A search of the NOV40a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 40C.

Table 40C. Geneseq Results for NOV40a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB10284	Human fetal placenta protein fragment AC175_2i - Homo sapiens, 173 aa. [WO200037630-A1, 29-JUN-2000]	1..148 1..173	148/173 (85%) 148/173 (85%)	5e-82
AAG03464	Human secreted protein, SEQ ID NO: 7545 - Homo sapiens, 91 aa. [EP1033401-A2, 06-SEP-2000]	1..76 1..91	76/91 (83%) 76/91 (83%)	5e-37
ABP41833	Human ovarian antigen HOPJF55, SEQ ID NO:2965 - Homo sapiens, 232 aa. [WO200200677-A1,	58..116 145..203	33/59 (55%) 48/59 (80%)	4e-17

	03-JAN-2002]			
AAU30569	Novel human secreted protein #1060 - Homo sapiens, 203 aa. [WO200179449-A2, 25-OCT-2001]	58..115 114..171	32/58 (55%) 47/58 (80%)	2e-16
AAY35324	Chlamydia pneumoniae transmembrane protein sequence - Chlamydia pneumoniae, 172 aa. [WO9927105-A2, 03-JUN-1999]	72..107 3..43	11/41 (26%) 19/41 (45%)	7.8

In a BLAST search of public sequence datbases, the NOV40a protein was found to have homology to the proteins shown in the BLASTP data in Table 40D.

Table 40D. Public BLASTP Results for NOV40a

Protein Accession Number	Protein/Organism/Length	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q13361	Microfibrillar-associated protein 5 precursor (MFAP-5) (Microfibril- associated glycoprotein 2) (MAGP-2) (MP25) - Homo sapiens (Human), 173 aa.	1..148 1..173	148/173 (85%) 148/173 (85%)	1e-81
Q28022	Microfibrillar-associated protein 5 precursor (MFAP-5) (Microfibril- associated glycoprotein 2) (MAGP-2) (MP25) - Bos taurus (Bovine), 170 aa.	1..148 1..170	118/170 (69%) 130/170 (76%)	2e-64
Q9QZJ6	Microfibrillar-associated protein 5 precursor (MFAP-5) (Microfibril- associated glycoprotein 2) (MAGP-2) - Mus musculus (Mouse), 164 aa.	1..148 1..164	118/168 (70%) 130/168 (77%)	7e-64
Q99PM0	Microfibril-associated glycoprotein 1 - Mus musculus (Mouse), 185 aa.	29..116 64..156	42/93 (45%) 58/93 (62%)	1e-17

P55002	Microfibrillar-associated protein 2 precursor (MFAP-2) (Microfibril- associated glycoprotein) (MAGP) (MAGP-1) - Mus musculus (Mouse), 183 aa.	29..116 62..154	42/93 (45%) 58/93 (62%)	1e-17

PFam analysis predicts that the NOV40a protein contains the domains shown in the Table 40E.

Table 40E. Domain Analysis of NOV40a			
Pfam Domain	**NOV40a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**

Example 41.

The NOV41 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 41A.

Table 41A. NOV41 Sequence Analysis			
	SEQ ID NO: 153	2518 bp	
NOV41a, CG151069-01 DNA Sequence	CAAAAGGGAACTTTATATGGAAAAGCTTCAAGAACATTTAATCAAAGCAAAAGCCTTTACCATAAAG AAGACGCTGGAGATCTATGTGCCCATCAAACAGTTCTTTTACAACCTCATCCACCCGGAGTATAGCG CCGTGACTGACGTGTATGTACTCATGTTCCTGGCTGACACTGTGGACTTCATCATCATTGTCTTCGG CTTTTGGGCCTTTGGGAAACACTCAGCAGCTGCAGACATCACCTCTTCACTGTCAGAGGACCAGGTC CCGGGGCCGTTTTTGGTGATGGTCCTCATTCAGTTTGGAACCATGGTGGTGGACCGAGCCCTCTACC TCAGGAAGACTGTACTGGGAAAGGTCATCTTCCAGGTCATTCTTGTGTTCGGAATTCACTTCTGGAT GTTCTTCATCTTACCTGGTGTGACTGAGAGGAAATTCAGCCAGAACCTGGTTGCCCAGCTTTGGTAC TTTGTGAAATGTGTTTACTTCGGGTTGTCTGCTTACCAGATCCGTTGTGGCTACCCAACGCGAGTCC TGGGGAACTTCCTCACCAAGAGCTACAATTACGTCAACCTCTTCTTATTCCAAGGGTTTCGCCTCGT GCCCTTTTTGACTGAGCTGAGGGCAGTGATGGACTGGGTGTGGACGGACACAACTTTGAGCCTGTCC AGCTGGATCTGTGTGGAGGACATCTATGCTCACATATTCATCCTGAAGTGTTGGCGGGAGTCGGAGA AGAGATACCCTCAGCCACGGGGCCAGAAGAAGAAGAAAGTGGTGAAGTATGGCATGGGAGGAATGAT CATCGTCCTGCTCATCTGCATTGTCTGGTTTCCTCTTCTCTTCATGTCTTTGATCAAATCTGTGGCT GGGGTCATCAACCAGCCCCTGGACGTCTCCGTCACAATTACCCTGGGAGGGTATCAGCCTATTTTCA CAATGAGTGCCCAACAAAGCCAGTTGAAAGTTATGGACCAGCAGAGCTTTAACAAATTTATACAAGC TTTTTCTAGGGACACCGGTGCTATGCAATTTCTGGAAAATTATGAAAAAGAAGACATAACAGTAGCA GAACTGGAAGGAAACTCAAATTCTTTGTGGACCATCAGCCCACCCAGTAAGCAGAAAATGATACACG AACTCCTGGACCCCAATAGTAGCTTCTCTGTTGTTTTTTCATGGAGTATTCAGAGAAACTTAAGTCT GGGTGCAAAATCGGAAATAGCAACAGATAAGCTTTCTTTTCCTCTTAAAAATATTACTCGAAAGAAT ATCGCTAAAATGATAGCAGGCAACAGCACAGAAAGTTCAAAAACACCAGTGACCATAGAAAAGATTT ATCCATATTATGTGAAAGCACCTAGTGATTCTAACTCAAAACCTATAAAGCAACTTTTATCTGAAAA TAATTTCATGGATATTACCATCATTTTGTCCAGAGACAATACAACTAAATATAACAGTGAGTGGTGG GTTCTCAACCTGACTGGAAACAGAATATACAATCCGAACTCTCAGGCCCTGGAACTGGTGGTCTTCA ATGACAAAGTCAGTCCCCCAAGTCTGGGGTTCCTGGCTGGCTATGGTATTATGGGATTATATGCTTC AGTTGTCCTTGTGATTGGGAAATTTGTCCGTGAATTCTTCAGTGGGATTTCTCACTCCATCATGTTT GAAGAGCTTCCAAATGTGGATCGAATTTTGAAGTTGTGCACAGATATTTTTTTAGTTCGAGAGACAG GAGAACTGGAGCTAGAAGAAGATCTCTATGCCAAATTAATATTCCTATATCGCTCACCAGAGACAAT GATCAAATGGACTAGAGAAAAAACAAATTGAAACCTTAGAACACAGACTGCAAATAATGTTAACATT TGAATTTTTTTTAAAAGCACAATATTCTCATAAGAGCTAAGCATTTCTAGTTCGACGGAAATGGTTT GTTTCTCTTCTGATAGGTAGACAAAAGGAGCTGATATCCTTCTGCAGTAAAAGCTACCTGGCAAGTT AAGGCACTGTTGAAAATGTTATTTGTAACTCCATTTCTCTGAAATCAGGGCTACTTGCTTTATGTTT TAGTCAACAGTGTCTCGCATTCTGATTGATCATGTGAAGGAATCATTTATGGGCCCCGTCCCTAAGA		

	GAAACAGAAGAGGAGTCAGAAAGAAAGATGCCTGTGTTTTCCTCTGTGGGGCCCGTGCACTTCCTGG AGAGATGCTACAATGCAATATACAGCGCTCCATCCCCACTGGGGAAGCTGCTGTGATGAGACTAGAT GAGCCTTCAACACACTCAGAAAATGCAACAGCAATAGGGGGCAGACAGCTCCTACCTGTGTTTCTAG GAGCAAAAGAGAGGGAACTAATTGCCCGTGAAGACGCCAGTGGAAGGATCAGCCTCATTCTAAGCAA AAACATAGTATTAGTGATACTCTTACTGCCTTATCTTAACCAAGGACTAATAGGATACCTTTCCATT AAACACCAGTGACTTCTCAGGAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 17		ORF Stop: TGA at 1838

	SEQ ID NO: 154	607 aa	MW at 69659.7kD
NOV41a, CG151069-01 Protein Sequence	MEKLQEHLIKAKAFTIKKTLEIYVPIKQFFYNLIHPEYSAVTDVYVLMFLADTVDFIIIVFGFWAFG KHSAAADITSSLSEDQVPGPFLVMVLIQFGTMVVDRALYLRKTVLGKVIFQVILVFGIHFWMFFILP GVTERKFSQNLVAQLWYFVKCVYFGLSAYQIRCGYPTRVLGNFLTKSYNYVNLFLFQGFRLVPFLTE LRAVMDWVWTDTTLSLSSWICVEDIYAHIFILKCWRESEKRYPQPRGQKKKKVVKYGMGGMIIVLLI CIVWFPLLFMSLIKSVAGVINQPLDVSVTITLGGYQPIFTMSAQQSQLKVMDQQSFNKFIQAFSRDT GAMQFLENYEKEDITVAELEGNSNSLWTISPPSKQKMIHELLDPNSSFSVVFSWSIQRNLSLGAKSE IATDKLSFPLKNITRKNIAKMIAGNSTESSKTPVTIEKIYPYYVKAPSDSNSKPIKQLLSENNFMDI TIILSRDNTTKYNSEWWVLNLTGNRIYNPNSQALELVVFNDKVSPPSLGFLAGYGIMGLYASVVLVI GKFVREFFSGISHSIMFEELPNVDRILKLCTDIFLVRETGELELEEDLYAKLIFLYRSPETMIKWTR EKTN		

Further analysis of the NOV41a protein yielded the following properties shown in Table 41B.

Table 41B. Protein Sequence Properties NOV41a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV41a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 41C.

Table 41C. Geneseq Results for NOV41a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY53635	A bone marrow secreted protein designated BMS53 - Homo sapiens, 466 aa. [WO9933979-A2, 08-JUL-1999]	141..605 2..465	290/471 (61%) 362/471 (76%)	e-169
ABB89128	Human polypeptide SEQ ID NO 1504 - Homo sapiens, 270 aa. [WO200190304-A2, 29-NOV-2001]	338..607 1..270	266/270 (98%) 266/270 (98%)	e-150

ABB63880	Drosophila melanogaster polypeptide SEQ ID NO 18432 - Drosophila melanogaster, 2771 aa. [WO200171042-A2, 27-SEP-2001]	28..605 2140..2740	233/607 (38%) 366/607 (59%)	e-125
AAB56086	Human secreted protein sequence encoded by gene 10 SEQ ID NO:180 - Homo sapiens, 379 aa. [WO200070042-A1, 23-NOV-2000]	246..605 20..378	201/366 (54%) 261/366 (70%)	e-109
ABB89513	Human polypeptide SEQ ID NO 1889 - Homo sapiens, 135 aa. [WO200190304-A2, 29-NOV-2001]	48..180 1..132	97/133 (72%) 105/133 (78%)	2e-48

In a BLAST search of public sequence datbases, the NOV41a protein was found to have homology to the proteins shown in the BLASTP data in Table 41D.

Table 41D. Public BLASTP Results for NOV41a

Protein Accession Number	Protein/Organism/Length	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H5I5	CDNA: FLJ23403 fis, clone HEP18857 - Homo sapiens (Human), 544 aa.	1..607 1..544	533/607 (87%) 537/607 (87%)	0.0
Q92508	Hypothetical protein KIAA0233 - Homo sapiens (Human), 2035 aa.	10..605 1440..2034	381/602 (63%) 467/602 (77%)	0.0
Q9VLS3	CG8486 protein - Drosophila melanogaster (Fruit fly), 2771 aa.	28..605 2140..2740	233/607 (38%) 366/607 (59%)	e-124
C88779	protein T20D3.9 [imported] - Caenorhabditis elegans, 1001 aa.	25..603 371..994	215/637 (33%) 337/637 (52%)	5e-96
Q9H5R4	CDNA: FLJ23144 fis, clone LNG09262 - Homo sapiens (Human), 150 aa.	423..572 1..150	150/150 (100%) 150/150 (100%)	2e-81

PFam analysis predicts that the NOV41a protein contains the domains shown in the Table 41E.

Table 41E. Domain Analysis of NOV41a			
Pfam Domain	**NOV41a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**

Example 42.

The NOV42 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 42A.

Table 42A. NOV42 Sequence Analysis			
	SEQ ID NO: 155	2035 bp	
NOV42a, CG151189-01 DNA Sequence	AGCGGGGCAGGTGGTGGCCGCCGGCCGGGCCCCGCCCTGGGGCCGCCTCCCCGCGGGTTCCGTTGGC TGTGGCGGCAGCTGACGCTTGTGGCGGCGGTGGCTTCGGGGTGGGCGTAAGATGGCGACAGCAGCGC AGGGACCCCTAAGCTTGCTGTGGGGCTGGCTGTGGAGCGAGCGCTTCTGGCTACCCGAGAACGTGAG CTGGGCTGATCTGGAGGGGCCGGCCGACGGCTACGGTTACCCCCGCGGCCGGCACATCCTCTCGGTG TTCCCGCTGGCGGCGGGCATCTTCTTCGTGAGGCTGCTCTTCGAGCGATTTATTGCCAAACCCTGTG CACTCCGTATTGGCATCGAGGACAGTGGTCCTTATCAGGCCCAACCCAATGCCATCCTTGAAAAGGT GTTCATATCTATTACCAAGTATCCTGATAAGAAAAGGCTGGAGGGCCTGTCAAAGCAGCTGGATTGG AATGTCCGAAAAATCCAATGCTGGTTTCGCCATCGGAGGAATCAGGACAAGCCCCCAACGCTTACTA AATTCTGTGAAAGCATGTGGAGATTCACATTTTATTTATGTATATTCTGCTATGGAATTAGATTTCT CTGGTCGTCACCTTGGTTCTGGGACATCCGACAGTGCTGGCATAACTATCCATTTCAGCCTCTTTCA AGTGGGCTTTATCACTATTATATCATGGAATTGGCCTTCTATTGGTCCCTTATGTTTTCTCAGTTTA CAGACATTAAAAGAAAGGACTTCCTGATCATGTTTGTGCATCACTTGGTCACCATTGGGCTTATCTC CTTCTCCTACATCAACAATATGGTTCGAGTGGGAACTCTGATCATGTGTCTACATGATGTCTCAGAC TTCTTGCTGGAGGCAGCCAAACTGGCCAATTATGCCAAGTATCAGCGGCTCTGTGACACCCTTTTTG TGATCTTCAGTGCTGTTTTTATGGTTACACGACTAGGAATCTATCCATTCTGGATTCTGAACACGAC CCTCTTTGAGAGTTGGGAGATAATCGGGCCTTATGCTTCATGGTGGCTCCTCAATGGCCTGCTGCTG ACCCTACAGCTTCTGCATGTCATCTGGTCCTACCTAATTGCACGGATTGCTTTGAAAGCCTTGATCA GGGGAAAGGTATCGAAGGATGATCGCAGTGATGTGGAGAGCAGCTCAGAGGAAGAAGATGTGACCAC CTGCACAAAAAGTCCCTGTGACAGTAGCTCCAGCAATGGTGCCAATCGGGTGAATGGTCACATGGGA GGCAGCTACTGGGCTGAAGAGTAAGGTGGTTGCTATAGGGACTTCAGCACACATGGACTTGTAGGGC CACTGGCAACATACTCCTCTTGGCCCTTCCCATATCTACTCTTCTGTGATTGGGAGACTGCAAGGCA CTGAGGAGTATCAAAGAAGCAAATATTTTCACTTTGAAAGAAAACTGCCATTTTGTATTTAATAGCC TCCAGGTTCTTTCAGTAATGTTATTTGCTCTGTGTGTTTTTGTGTGTTTGTTGATGTGCGTTTGTGC ATATGCGTGAGTTTCATTGCCGGGGTTGGGGCACAATTGTGGACTGGGGCCATGAGGCCTTCCCTGG TCCCCACTGAACCCACCTTAGTTCCACATTTGGCTGCATCTTGAATTATGCCGACTCCAGACTTCTC CTCCTTTTTTGCCCTTGGCTCTTGACACTCTAAACCCCTGGACCATCTGAATGGAGCAGCCAAGTTC AGTCCCACATTTCTGTACTGTTCCTCTTTCACAGCTGGAATATGTCACATGATGAAGTTGTATAGAA ACAGAACCATGGATGGATGGCCAGGATTGCCGTGGTCCCTAGCTAGATCCCCTTCCTATCAATCACC TGATAGCAACAGGGACAGCTGCCAATACCCTGCTCTTTACTCAATGGTACCCAGGGAGGGAGCATGG GAAGAGGGTGAGCTGAGGGCTGGAGGAGGGCAACAGCCACTGGGTGAGCTGTTCACGGTCTTATACT ATTGTTTGTGATTAAAAGTGCTTCA		
	ORF Start: ATG at 119		ORF Stop: TAA at 1295

	SEQ ID NO: 156	392 aa	MW at 45804.6kD
NOV42a, CG151189-01 Protein Sequence	MATAAQGPLSLLWGWLWSERFWLPENVSWADLEGPADGYGYPRGRHILSVFPLAAGIFFVRLLFERF IAKPCALRIGIEDSGPYQAQPNAILEKVFISITKYPDKKRLEGLSKQLDWNVRKIQCWFRHRRNQDK PPTLTKFCESMWRFTFYLCIFCYGIRFLWSSPWFWDIRQCWHNYPFQPLSSGLYHYYIMELAFYWSL MFSQFTDIKRKDFLIMFVHHLVTIGLISFSYINNMVRVGTLIMCLHDVSDFLLEAAKLANYAKYQRL CDTLFVIFSAVFMVTRLGIYPFWILNTTLFESWEIIGPYASWWLLNGLLLTLQLLHVIWSYLIARIA LKALIRGKVSKDDRSDVESSSEEEDVTTCTKSPCDSSSNGANRVNGHMGGSYWAEE		

Further analysis of the NOV42a protein yielded the following properties shown in Table 42B.

Table 42B. Protein Sequence Properties NOV42a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3104 probability located in mitochondrial intermembrane space; 0.3000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 31 and 32

A search of the NOV42a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 42C.

Table 42C. Geneseq Results for NOV42a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU00782	Human apoptosis protein, APOP-2 - Homo sapiens, 311 aa. [WO200118042-A2, 15-MAR-2001]	100..392 19..311	291/293 (99%) 293/293 (99%)	e-179
ABB90335	Human polypeptide SEQ ID NO 2711 - Homo sapiens, 296 aa. [WO200190304-A2, 29-NOV-2001]	15..293 7..284	183/279 (65%) 230/279 (81%)	e-116
AAB93884	Human protein sequence SEQ ID NO:13813 - Homo sapiens, 394 aa. [EP1074617-A2, 07-FEB-2001]	15..361 8..353	160/347 (46%) 218/347 (62%)	4e-90
ABB90167	Human polypeptide SEQ ID NO 2543 - Homo sapiens, 394 aa. [WO200190304-A2, 29-NOV-2001]	15..360 8..352	159/346 (45%) 217/346 (61%)	1e-89
AAM78909	Human protein SEQ ID NO 1571 - Homo sapiens, 394 aa. [WO200157190-A2, 09-AUG-2001]	15..360 8..352	159/346 (45%) 217/346 (61%)	1e-89

In a BLAST search of public sequence datbases, the NOV42a protein was found to have homology to the proteins shown in the BLASTP data in Table 42D.

Table 42D. Public BLASTP Results for NOV42a				
Protein Accession Number	**Protein/Organism/Length**	**NOV42a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAH32565	Similar to RIKEN cDNA 2310081H14 gene - Homo sapiens (Human), 392 aa.	1..392 1..392	391/392 (99%) 391/392 (99%)	0.0
Q924Z3	TRH4 - Mus musculus (Mouse), 414 aa.	1..392 1..392	301/392 (76%) 339/392 (85%)	0.0
Q9D6K9	2310081H14Rik protein - Mus musculus (Mouse), 414 aa.	1..392 1..392	301/392 (76%) 339/392 (85%)	0.0
Q8QGA3	TRH4 - Xenopus laevis (African clawed frog), 382 aa.	9..392 1..382	288/385 (74%) 326/385 (83%)	e-179
Q90YY6	Trh1 - Brachydanio rerio (Zebrafish) (Zebra danio), 406 aa.	15..360 11..358	166/348 (47%) 224/348 (63%)	4e-92

PFam analysis predicts that the NOV42a protein contains the domains shown in the Table 42E.

Table 42E. Domain Analysis of NOV42a			
Pfam Domain	**NOV42a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
homeobox	92..135	16/44 (36%) 28/44 (64%)	0.029

Example 43.

The NOV43 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 43A.

Table 43A. NOV43 Sequence Analysis			
	SEQ ID NO: 157	1845 bp	
NOV43a, CG151801-01 DNA Sequence	GTGTGAAAATCACAAATGTCAAATGATGGAAGATCCAGGAATCGGGACAGGCGCTACGATGAGGTCC CAAGCGACCTGCCCTATCAAGATACCACCATAAGAACCCACCCAACTCTTCATGACAGTGAGCGGGC AGTGAGCGCTGATCCCTTGCCACCACCCCCTCTCCCATTACAGCCACCATTCGGCCCAGACTTCTAC TCAAGTGACACAGAAGAACCAGCTATAGCGCCAGATCTCAAACCAGTAAGGCGCTTTGTCCCTGACT CCTGGAAGAACTTTTTCAGAGGGAAGAAAAAGGACCCCGAATGGGATAAGCCGGTGTCTGATATCAG GTACATCTCCGATGGAGTGGAGTGTTCACCACCAGCCTCTCCAGCAAGACCAAACCACCGTTCGCCC		

	CTCAACTCCTGCAAAGATCCCTACGGCGGGTCAGAAGGAACCTTTAGTTCCCGGAAAGAGGCTGACG CAGTGTTTCCCCGGGATCCCTATGGATCTCTAGACCGACACACACAAACAGTTCGAACATACAGTGA GAAGGTGGAGGAGTATAACCTGAGATACTCCTACATGAAGTCGTGGCAGGCCTGCTGAGAATACTG GGTGTGGTGGAGCTGCTTTTGGGGGCCGGTGTCTTTGCTTGTGTCACAGCTTACATTCACAAGGACA GTGAGTGGTACAACTTGTTTGGATATTCACAACCGTATGGCATGGGAGGCGTTGGTGGATTGGGCAG TATGTATGGGGGCTATTACTACACTGGCCCTAAGACCCCTTTTGTACTCGTGGTTGCTGGATTAGCT TGGATCACCACCATTATTATTCTGGTTCTTGGCATGTCCATGTATTACCGGACCATTCTTCTGGACT CTAATTGGTGGCCCCTAACTGAATTTGGAATTAACGTTGCCTTGTTTATTTTGTATATGGCCGCAGC CATAGTCTATGTGAATGATACCAACCGAGGTGGCCTCTGCTACTATCCGTTATTTAATACACCAGTG AATGCAGTGTTCTGCCGGGTAGAAGGAGGACAGATAGCTGCAATGATCTTCCTGTTTGTCACCATGA TAGTTTATCTCATTAGTGCTTTGGTTTGCCTAAAGTTATGGAGGCATGAGGCAGCTCGGAGACATAG AGAATATATGGAACAACAGGAGGTAAGTGATATAAATGAGCCATCATTGTCATCGAAAAGGAAAATG TGTGAAATGGCCACCAGTGGTGACAGACAAAGAGACTCAGAAGTTAATTTCAAGGAACTGAGAACAG CAAAAATGAAACCTGAACTACTGAGTGGACACATCCCCCCAGGCCACATTCCTAAACCTATCGTGAT GCCCGACTATGTGGCGAAATACCCTGTGATTCAGACAGATGATGAGCGAGAACGCTATAAAGCTGTG TTCCAAGACCAGTTTTCAGAGTACAAAGAGCTGTCTGCAGAAGTTCAGGCTGTCCTGAGGAAGTTTG ATGAGCTGGATGCAGTGATGAGCAGATTGCCACATCATTCGGAAAGCCGACAGGAACATGAGAGAAT TTCAAGAATCCATGAAGAGTTTAAGAAAAAAAAGAATGATCCTACATTTCTGGAAAAAAAAGAACGC TGTGATTACCTAAAGAATAAACTTTCTCACATAAAGCAAAGAATTCAAGAATATGATAAAGTAATGA ATTGGATGTACAAGGTTATTCTTAACGCTTATTTGAAACCACTTTATTTTTTTATTTTATTTTATT TTTTTGAGATGAAGTCTCGCTCTGTTACCCAGGCTGGAATGCAGTGGCACAATCTCGGCTCACTGCA ACCTCCACCTCCCGGGTTCAAGCAATTCTCCTGTTC		
	ORF Start: ATG at 16		ORF Stop: TAA at 1699

	SEQ ID NO: 158	561 aa	MW at 64468.7kD
NOV43a, CG151801-01 Protein Sequence	MSNDGRSRNRDRRYDEVPSDLPYQDTTIRTHPTLHDSERAVSADPLPPPPLPLQPPFGPDFYSSDTE EPAIAPDLKPVRRFVPDSWKNFFRGKKKDPEWDKPVSDIRYISDGVECSPPASPARPNHRSPLNSCK DPYGGSEGTFSSRKEADAVFPRDPYGSLDRHTQTVRTYSEKVEEYNLRYSYMKSWAGLLRILGVVEL LLGAGVFACVTAYIHKDSEWYNLFGYSQPYGMGGVGGLGSMYGGYYYTGPKTPFVLVVAGLAWITTI IILVLGMSMYYRTILLDSNWWPLTEFGINVALFILYMAAAIVYVNDTNRGGLCYYPLFNTPVNAVFC RVEGGQIAAMIFLFVTMIVYLISALVCLKLWRHEAARRHREYMEQQEVSDINEPSLSSKRKMCEMAT SGDRQRDSEVNFKELRTAKMKPELLSGHIPPGHIPKPIVMPDYVAKYPVIQTDDERERYKAVFQDQF SEYKELSAEVQAVLRKFDELDAVMSRLPHHSESRQEHERISRIHEEFKKKKNDPTFLEKKERCDYLK NKLSHIKQRIQEYDKVMNWDVQGYS		

Further analysis of the NOV43a protein yielded the following properties shown in Table 43B.

Table 43B. Protein Sequence Properties NOV43a	
PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV43a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 43C.

Table 43C. Geneseq Results for NOV43a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

ABG14568	Novel human diagnostic protein #14559 - Homo sapiens, 363 aa. [WO200175067-A2, 11-OCT-2001]	1..332 1..333	325/333 (97%) 326/333 (97%)	0.0
AAB82940	Human androgen receptor trapped protein 5 (ART5) - Homo sapiens, 264 aa. [WO200172332-A1, 04-OCT-2001]	392..550 94..253	63/161 (39%) 98/161 (60%)	2e-26
AAB56085	Human secreted protein sequence encoded by gene 9 SEQ ID NO:179 - Homo sapiens, 264 aa. [WO200070042-A1, 23-NOV-2000]	392..550 94..253	63/161 (39%) 98/161 (60%)	2e-26
AAW76212	Human ELL2 protein - Homo sapiens, 640 aa. [WO9837194-A1, 27-AUG-1998]	371..551 466..633	60/184 (32%) 100/184 (53%)	6e-19
AAB57048	Human prostate cancer antigen protein sequence SEQ ID NO:1626 - Homo sapiens, 677 aa. [WO200055174-A1, 21-SEP-2000]	371..551 503..670	60/184 (32%) 99/184 (53%)	1e-18

In a BLAST search of public sequence datbases, the NOV43a protein was found to have homology to the proteins shown in the BLASTP data in Table 43D.

Table 43D. Public BLASTP Results for NOV43a

Protein Accession Number	Protein/Organism/Length	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH33689	Similar to hypothetical protein FLJ30532 - Homo sapiens (Human), 558 aa.	1..561 1..558	557/561 (99%) 557/561 (99%)	0.0
Q96NM9	CDNA FLJ30532 fis, clone BRAWH2001129, weakly similar to occludin - Homo sapiens (Human), 457 aa.	1..433 1..430	429/433 (99%) 429/433 (99%)	0.0
Q99LE8	Hypothetical 50.4 kDa protein - Mus musculus (Mouse), 436 aa (fragment).	121..560 1..435	386/441 (87%) 409/441 (92%)	0.0

Q9H607	CDNA: FLJ22709 fis, clone HSI13338 - Homo sapiens (Human), 264 aa.	392..550 94..253	63/161 (39%) 98/161 (60%)	5e-26
Q8VCR9	Similar to RIKEN cDNA 9430098E02 gene - Mus musculus (Mouse), 219 aa.	437..550 94..206	49/114 (42%) 73/114 (63%)	2e-20

PFam analysis predicts that the NOV43a protein contains the domains shown in the Table 43E.

Table 43E. Domain Analysis of NOV43a			
Pfam Domain	**NOV43a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
Occludin	444..553	33/110 (30%) 56/110 (51%)	6.2e-09

Example 44.

The NOV44 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 44A.

Table 44A. NOV44 Sequence Analysis			
	SEQ ID NO: 159	1112 bp	
NOV44a, CG165961-01 DNA Sequence	TGAGGCGAGTGAAGTGGACTCTGAGGGCTACCGCTACCGCCACTGCTGCGGCAGGGGCGTGGAGGGC AGAGGGCCGCGGAGGCCGCAGTTGCAAACATGGCTCAGAGCAGAGACGGCGGAAACCCGTTCGCCGA GCCCAGCGAGCTTGACAACCCCTTTCAGGACCCAGCTGTGATCCAGCACCGACCCAGCCGGCAGTAT GCCACGCTTGACGTCTACAACCCTTTTGAGACCCGGGAGGCCTCAGCTGCAGCAGCCACAGCTGAGC TGCTGAAGAAACAGGAGGAGCTCAACCGGAAGGCAGAGGAGTTGGACCGAAGGGAGCGAGAGCTGCA GCATGCTGCCCTGGGAGGCACAGCTACTCGACAGAACAATTGGCCCCCTCTACCTTCTTTTTGTCCA GTTCAGCCCTGCTTTTTCCAGGACATCTCCATGGAGATCCCCCAAGAATTTCAGAAGACTGTATCCA CCATGTACTACCTCTGGATGTGCAGCACGCTGGCTCTTCTCCTGAACTTCCTCGCCTGCCTGGCCAG CTTCTGTGTGGAAACCAACAATGGCGCAGGCTTTGGGCTTTCTATCCTCTGGGTCCTCCTTTTCACT CCCTGCTCCTTTGTCTGCTGGTACCGCCCCATGTATAAGGCTTTCCGGAGTGACAGTTCATTCAATT TCTTCGTTTTCTTCTTCATTTTCTTCGTCCAGGATGTGCTCTTTGTCCTCCAGGCCATTGGTATCCC AGGTTGGGGATTCAGTGGCTGGATCTCTGCTCTGGTGGTGCCGAAGGGCAACACAGCAGTATCCGTG CTCATGCTGCTGGTCGCCCTGCTCTTCACTGGCATTGCTGTGCTAGGAATTGTCATGCTGAAACGGA TCCACTCCTTATACCGCCGCACAGGTGCCAGCTTTCAGAAGGCCCAGCAAGAATTTGCTGCTGGTGT CTTCTCCAACCCTGCGGTGCGAACCGCAGCTGCCAATGCAGCCGCTGGGGCTGCTGAAAATGCCTTC CGGGCCCCGTGACCCCTGACTGGGATGCCCTGGCCCTGCTACTTGAGGGAGCTGACTTAGCTCCCGT CCCTAAGGTCTCTGGGACTTGGAGAGACATCACTAACTGA		
	ORF Start: ATG at 97		ORF Stop: TGA at 1015

	SEQ ID NO: 160	306 aa	MW at 33990.7kD
NOV44a, CG165961-01 Protein Sequence	MAQSRDGGNPFAEPSELDNPFQDPAVIQHRPSRQYATLDVYNPFETREASAAAATAELLKKQEELNR KAEELDRRERELQHAALGGTATRQNNWPPLPSFCPVQPCFFQDISMEIPQEFQKTVSTMYYLWMCST LALLLNFLACLASFCVETNNGAGFGLSILWVLLFTPCSFVCWYRPMYKAFRSDSSFNFFVFFFIFFV QDVLFVLQAIGIPGWGFSGWISALVVPKGNTAVSVLMLLVALLFTGIAVLGIVMLKRIHSLYRRTGA SFQKAQQEFAAGVFSNPAVRTAAANAAAGAAENAFRAP		

	SEQ ID NO: 161	1310 bp	
NOV44b, CG165961-02 DNA Sequence	TGAGGCGAGTGAAGTGGACTCTGAGGGCTACCGCTACCGCCACTGCTGCGGCAGGGGCGTGGAGGGC AGAGGGCCGCGGAGGCCGCAGTTGCAAACATGGCTCAGAGCAGAGACGGCGGAAACCCGTTCGCCGA GCCCAGCGAGCTTGACAACCCCTTTCAGGACCCAGCTGTGATCCAGCACCGACCCAGCCGGCAGTAT GCCACGCTTGACGTCTACAACCCTTTTGAGACCCGGGAGGCCTCAGCTGCAGCAGCCACAGCTGAGC TGCTGAAGAAACAGGAGGAGCTCAACCGGAAGGCAGAGGAGTTGGACCGAAGGGAGCGAGAGCTGCA GCATGCTGCCCTGGGAGGCACAGCTACTCGACAGAACAATTGGCCCCCTCTACCTTCTTTTTGTCCA GTTCAGCCCTGCTTTTTCCAGGACATCTCCATGGAGATCCCCCAAGAATTTCAGAAGACTGTATCCA CCATGTACTACCTCTGGATGTGCAGCACGCTGGCTCTTCTCCTGAACTTCCTCGCCTGCCTGGCCAG CTTCTGTGTGGAAACCAACAATGGCGCAGGCTTTGGGCTTTCTATCCTCTGGGTCCTCCTTTTCACT CCCTGCTCCTTTGTCTGCTGGTACCGCCCCATGTATAAGGCTTTCCGGAGTGACAGTTCATTCAATT TCTTCGTTTTCTTCTTCATTTTCTTCGTCCAGGATGTGCTCTTTGTCCTCCAGGCCATTGGTATCCC AGGTTGGGGATTCAGTGGCTGGATCTCTGCTCTGGTGGTGCCGAAGGGCAACACAGCAGTATCCGTG CTCATGCTGCTGGTCGCCCTGCTCTTCACTGGCATTGCTGTGCTAGGAATTGTCATGCTGAAACGGA TCCACTCCTTATACCGCCGCACAGGTGCCAGCTTTCAGAAGGCCCAGCAAGAATTTGCTGCTGGTGT CTTCTCCAACCCTGCGGTGCGAACCGCAGCTGCCAATGCAGCCGCTGGGGCTGCTGAAAATGCCTTC CGGGCCCCGTGACCCCTGACTGGGATGCCCTGGCCCTGCTACTTGAGGGAGCTGACTTAGCTCCCGG CCCTAAGGTCTCTGGGACTTGGAGAGACATCACTAACTGATGGCTCCTCCGTAGTGCTCCCAATCCT ATGGCCATGACTGCTGAACCTGACAGGCGTGTGGGGAGTTCACTGTGACCTAGTCCCCCCATCAGGC CACACTGCTGCCACCTCTCACACGCCCCAACCCAGCTTCCCTCTGCTGTGCCACGGCTGTTGCTTCG GTTATTTAAATAAAAAGAAAGTGGAACTGGAACTGAC		
	ORF Start: ATG at 97		ORF Stop: TGA at 1015

	SEQ ID NO: 162	306 aa	MW at 33990.7kD
NOV44b, CG165961-02 Protein Sequence	MAQSRDGGNPFAEPSELDNPFQDPAVIQHRPSRQYATLDVYNPFETREASAAAATAELLKKQEELNR KAEELDRRERELQHAALGGTATRQNNWPPLPSFCPVQPCFFQDISMEIPQEFQKTVSTMYYLWMCST LALLLNFLACLASFCVETNNGAGFGLSILWVLLFTPCSFVCWYRPMYKAFRSDSSFNFFVFFFIFFV QDVLFVLQAIGIPGWGFSGWISALVVPKGNTAVSVLMLLVALLFTGIAVLGIVMLKRIHSLYRRTGA SFQKAQQEFAAGVFSNPAVRTAAANAAAGAAENAFRAP		

	SEQ ID NO: 163	1235 bp	
NOV44c, CG165961-03 DNA Sequence	TGAGGCGAGTGAAGTGGACTCTGAGGGCTACCGCTACCGCCACTGCTGCGGCAGGGGCGTGGAGGGC AGAGGGCCGCGGAGGCCGCAGTTGCGAACATGGCTCAGAGCAGAGACGGCGGAAACCCGTTCGCCGA GCCCAGCGAGCTTGACAACCCCTTTCAGGACCCACCTGTGATCCAGCACCGACCCAGCCGGCAGTAT GCCACGCTTGACGTCTACAACCCTTTTGAGACCCGGGAGCCACCACCAGCCTATGAGCCTCCAGCCC CTGCCCCATTGCCTCCACCCTCAGCTCCCTCCTTGCAGCCCTCGAGAAAGCTCAGCCCCACAGAACC TAAGAACTATGGCTCATACAGCACTCAGGCCTCAGCTGCAGCAGCCACAGCTGAGCTGCTGAAGAAA CAGGAGGAGCTCAACCGGAAGGCAGAGGAGTTGGACCGAAGGGAGCGAGAGCTGCAGCATGCTGCCC TGGGGGGCACAGCTACTCGACAGAACAATTGGCCCCCTCTACCTTCTTTTTGTCCAGTTCAGCCCTG CTTTTTCCAGGACATCTCCATGGAGATCCCCCAAGAATTTCAGAAGACTGTATCCACCATGTACTAC CTCTGGATGTGCAGCACGCTGGCTCTTCTCCTGAACTTCCTCGCCTGCCTGGCCAGCTTCTGTGTGG AAACCAACAATGGCGCAGGCTTTGGGCTTTCTATCCTCTGGGTCCTCCTTTTCACTCCCTGCTCCTT TGTCTGCTGGTACCGCCCCATGTATAAGGCTTTCCGGAGTGACAGTTCATTCAATTTCTTCGTTTTC TTCTTCATTTTCTTCGTCCAGGATGTGCTCTTTGTCCTCCAGGCCATTGGTATCCCAGGTTGGGGAT TCAGTGGCTGGATCTCTGCTCTGGTGGTGCCGAAGGGCAACACAGCAGTATCCGTGCTCATGCTGCT GGTCGCCCTGCTCTTCACTGGCATTGCTGTGCTAGGAATTGTCATGCTGAAACGGATCCACTCCTTA TACCGCCGCACAGGTGCCAGCTTTCAGAAGGCCCAGCAAGAATTTGCTGCTGGTGTCTTCTCCAACC CTGCGGTGCGAACCGCAGCTGCCAATGCAGCCGCTGGGGCTGCTGAAAATGCCTTCCGGGCCCCGTG ACCCCTGACTGGGATGCCCTGGCCCTGCTACTTGAGGGAGCTGACTTAGCTCCCGTCCCTAAGGTCT CTGGGACTTGGAGAGACATCACTAACTGA		
	ORF Start: ATG at 97		ORF Stop: TGA at 1138

	SEQ ID NO: 164	347 aa	MW at 38312.5kD
NOV44c, CG165961-03 Protein Sequence	MAQSRDGGNPFAEPSELDNPFQDPPVIQHRPSRQYATLDVYNPFETREPPPAYEPPAPAPLPPPSAP SLQPSRKLSPTEPKNYGSYSTQASAAAATAELLKKQEELNRKAEELDRRERELQHAALGGTATRQNN WPPLPSFCPVQPCFFQDISMEIPQEFQKTVSTMYYLWMCSTLALLLNFLACLASFCVETNNGAGFGL SILWVLLFTPCSFVCWYRPMYKAFRSDSSFNFFVFFFIFFVQDVLFVLQAIGIPGWGFSGWISALVV PKGNTAVSVLMLLVALLFTGIAVLGIVMLKRIHSLYRRTGASFQKAQQEFAAGVFSNPAVRTAAANA AAGAAENAFRAP		

	SEQ ID NO: 165	1543 bp	
NOV44d, CG165961-04 DNA Sequence	CGGCCGCGTCGACGGACTCTGAGGGCTACCGCTACCGCCACTGCTGCGGCAGGGGCGTGGAGGGCAG AGGGCCGCGGAGGCCGCAGTTGCAAACATGGCTCAGAGCAGAGACGGCGGAAACCCGTTCGCCGAGC CCAGCGAGCTTGACAACCCCTTTCAGCCACCACCAGCCTATGAGCCTCCAGCCCCTGCCCCATTGCC TCCACCCTCAGCTCCCTCCTTGCAGCCCTCGAGAAAGCTCAGCCCCACAGAACCTAAGAACTATGGC TCATACAGCACTCAGGCCTCAGCTGCAGCAGCCACAGCTGAGCTGCTGAAGAAACAGGAGGAGCTCA ACCGGAAGGCAGAGGAGTTGGACCGAAGGGAGCGAGAGCTGCAGCATGCTGCCCTGGGGGGCACAGC TACTCGACAGAACAATTGGCCCCCTCTACCTTCTTTTTGTCCAGTTCAGCCCTGCTTTTTCCAGGAC ATCTCCATGGAGATCCCCCAAGAATTTCAGAAGACTGTATCCACCATGTACTACCTCTGGATGTGCA GCACGCTGGCTCTTCTCCTGAACTTCCTCGCCTGCCTGGCCAGCTTCTGTGTGGAAACCAACAATGG CGCAGGCTTTGGGCTTTCTATCCTCTGGGTCCTCCTTTTCACTCCCTGCTCCTTTGTCTGCTGGTAC CGCCCCATGTATAAGGCTTTCCGGAGTGACAGTTCATTCAATTTCTTCGTTTTCTTCTTCATTTTCT TCGTCCAGGATGTGCTCTTTGTCCTCCAGGCCATTGGTATCCCAGGTTGGGGATTCAGTGGCTGGAT CTCTGCTCTGGTGGTGCCGAAGGGCAACACAGCAGTATCCGTGCTCATGCTGCTGGTCGCCCTGCTC TTCACTGGCATTGCTGTGCTAGGAATTGTCATGCTGAAACGGATCCACTCCTTATACCGCCGCACAG GTGCCAGCTTTCAGAAGGCCCAGCAAGAATTTGCTGCTGGTGTCTTCTCCAACCCTGCGGTGCGAAC CGCAGCTGCCAATGCAGCCGCTGGGGCTGCTGAAAATGCCTTCCGGGCCCCG**TGA**CCCCTGACTGGG ATGCCCTGGCCCTGCTACTTGAGGGAGCTGACTTAGCTCCCGTCCCTAAGGTCTCTGGACTTGGAG AGACATCACTAACTGATGGCTCCTCCGTAGTGCTCCCAATCCTATGGCCATGACTGCTGAACCTGAC AGGCGTGTGGGGAGTTCACTGTGACCTAGTCCCCCCATCAGGCCACACTGCTGCCACCTCTCACACG CCCCAACCCAGCTTCCCTCTGCTGTGCCACGGCTGTTGCTTCGGTTATTTAAATAAAAAGAAAGTGG AACTGGAACTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAACTATAATTTTTTTTTTTTTTTTTTTTTTTTACCCCCCCCGCTTTTTTTTTTTTTTTT TTTTTTTCCCCCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGTTTTTTTTTTTTTTTTTTTCCC CC		
	ORF Start: ATG at 95		ORF Stop: TGA at 1058

	SEQ ID NO: 166	321 aa	MW at 35201.1kD
NOV44d, CG165961-04 Protein Sequence	MAQSRDGGNPFAEPSELDNPFQPPPAYEPPAPAPLPPPSAPSLQPSRKLSPTEPKNYGSYSTQASAA AATAELLKKQEELNRKAEELDRRERELQHAALGGTATRQNNWPPLPSFCPVQPCFFQDISMEIPQEF QKTVSTMYYLWMCSTLALLLNFLACLASFCVETNNGAGFGLSILWVLLFTPCSFVCWYRPMYKAFRS DSSFNFFVFFFIFFVQDVLFVLQAIGIPGWGFSGWISALVVPKGNTAVSVLMLLVALLFTGIAVLGI VMLKRIHSLYRRTGASFQKAQQEFAAGVFSNPAVRTAAANAAAGAAENAFRAP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 44B.

Table 44B. Comparison of NOV44a against NOV44b through NOV44d.		
Protein Sequence	**NOV44a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV44b	1..285 1..285	223/285 (78%) 223/285 (78%)
NOV44c	1..285 1..326	228/326 (69%) 229/326 (69%)
NOV44d	1..285 1..300	201/300 (67%) 203/300 (67%)

Further analysis of the NOV44a protein yielded the following properties shown in Table 44C.

Table 44C. Protein Sequence Properties NOV44a

PSort analysis:	0.6000 probability located in plasma membrane; 0.4000 probability located in Golgi body; 0.3000 probability located in endoplasmic reticulum (membrane); 0.0300 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV44a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 44D.

Table 44D. Geneseq Results for NOV44a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV44a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB62698	Human membrane recycling protein (HMRP)-1 - Homo sapiens, 347 aa. [US6235715-B1, 22-MAY-2001]	1..306 1..347	306/347 (88%) 306/347 (88%)	e-173
AAY30521	A human membrane recycling protein designated HMRP-1 - Homo sapiens, 347 aa. [US5962263-A, 05-OCT-1999]	1..306 1..347	306/347 (88%) 306/347 (88%)	e-173
AAB62700	Rat SCAMP 37 protein - Rattus sp, 338 aa. [US6235715-B1, 22-MAY-2001]	9..304 7..334	180/333 (54%) 217/333 (65%)	4e-92
ABG61921	Prostate cancer-associated protein #122 - Mammalia, 338 aa. [WO200230268-A2, 18-APR-2002]	9..304 7..334	178/333 (53%) 215/333 (64%)	3e-90
AAB62699	Human membrane recycling protein (HMRP)-2 - Homo sapiens, 329 aa. [US6235715-B1, 22-MAY-2001]	9..298 7..325	164/322 (50%) 220/322 (67%)	1e-87

In a BLAST search of public sequence datbases, the NOV44a protein was found to have homology to the proteins shown in the BLASTP data in Table 44E.

Table 44E. Public BLASTP Results for NOV44a

Protein Accession Number	Protein/Organism/Length	NOV44a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O14828	Secretory carrier-associated membrane protein 3 - Homo sapiens (Human), 347 aa.	1..306 1..347	302/347 (87%) 302/347 (87%)	e-170
T08826	secretory carrier membrane protein homolog propin1 - human, 347 aa.	1..306 1..347	301/347 (86%) 301/347 (86%)	e-169
Q99M48	Similar to secretory carrier membrane protein 3 - Mus musculus (Mouse), 350 aa.	1..306 1..350	277/350 (79%) 289/350 (82%)	e-156
Q9ERM9	Secretory carrier membrane protein 3 - Mus musculus (Mouse), 349 aa.	1..306 1..349	276/349 (79%) 288/349 (82%)	e-155
O35609	Secretory carrier-associated membrane protein 3 - Mus musculus (Mouse), 349 aa.	1..306 1..349	274/349 (78%) 286/349 (81%)	e-154

PFam analysis predicts that the NOV44a protein contains the domains shown in the Table 44F.

Table 44F. Domain Analysis of NOV44a			
Pfam Domain	NOV44a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 45.

The NOV45 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 45A.

Table 45A. NOV45 Sequence Analysis			
	SEQ ID NO: 167	1356 bp	
NOV45a, CG171681-01 DNA Sequence	CTGCGCTGCCGAGGCGAGCTAAGCGCCCGCTCGCCATGGGGAGCCCCGCACATCGGCCCGCGCTGCT GCTGCTGCTGCCGCCTCTGCTGCTGCTGCTGCTGCTGCGCGTCCCGCCCAGCCGCAGCTTCCCAGAT ATGGAACCTCCTAGAATCAAGTGCCCAAGTGTGAAGGAACGCATTGCAGAACCCAACAAACTGACAG TCCGGGTGTCCTGGGAGACACCCGAAGGAAGAGACACAGCAGATGGAATTCTTACTGATGTCATTCT AAAAGGCCTCCCCCCAGGCTCCAACTTTCCAGAAGGAGACCACAAGATCCAGTACACAGTCTATGAC AGAGCTGAGAATAAGGGCACTTGCAAATTTCGAGTTAAAGTAAGAGTCAAACGCTGTGGCAAACTCA ATGCCCCAGAGAATGGTTACATGAAGTGCTCCAGCGACGGTGATAATTATGGAGCCACCTGTGAGTT CTCCTGCATCGGCGGCTATGAGCTCCAGGGTAGCCCTGCCCGAGTATGTCAATCCAACCTGGCTTGG TCTGGCACGGAGCCCACCTGTGCAGCCATGAACGTCAATGTGGGTGTCAGAACGGCAGCTGCACTTC TGGATCAGTTTTATGAGAAAAGGAGACTCCTCATTGTGTCCACACCCACAGCCCGAAACCTCCTTTA		

	CCGGCTCCAGCTAGGAATGCTGCAGCAAGCACAGTGTGGCCTTGATCTTCGACACATCACCGTGGTG GAGCTGGTGGGTGTGTTCCCGACTCTCATTGGCAGGATAGGAGCAAAGATTATGCCTCCAGCCCTAG CGCTGCAGCTCAGGCTGTTGCTGCGAATCCCACTCTACTCCTTCAGTATGGTGCTAGTGGATAAGCA TGGCATGGACAAAGAGCGCTATGTCTCCCTGGTGATGCCTGTGGCCCTGTTCAACCTGATTGACACT TTTCCCTTGAGAAAAGAAGAGATGGTCCTACAAGCCGAAATGAGCCAGACCTGTAACACCTGACATG ATGGTTCCTCTCTTGGCAATTCCTCTTCATTGTCTACATAGTGACATGCACACGGGAAAGCCTTAAA AATATCCTTGATGTACAGATTTTATTTGTAATTTTAAAAGTCTATTTTATTATGAGCTTTCTTTGCA CTTAAAAATTAGCATGCTGCTTTTTGTACTTGGAAGTGTTTCAAAAAATTATATGACCATATTTACT CTTTCTAACCTTTCTTTACTCCATCATGGCTGGTTGATTTGTAGAGAAATTAGAACCCATAACCATA CACAGGCTATCAACATGTTATTCAATGTGACACCTAACTCTTTTCTATTTTGTTTTTTAAGTAAGAC TTTTATTAATAAAACG		
	ORF Start: ATG at 36		ORF Stop: TGA at 999

	SEQ ID NO: 168	321 aa	MW at 35636.4kD
NOV45a, CG171681-01 Protein Sequence	MGSPAHRPALLLLLPPLLLLLLLRVPPSRSFPDMEPPRIKCPSVKERIAEPNKLTVRVSWETPEGRD TADGILTDVILKGLPPGSNFPEGDHKIQYTVYDRAENKGTCKFRVKVRVKRCGKLNAPENGYMKCSS DGDNYGATCEFSCIGGYELQGSPARVCQSNLAWSGTEPTCAAMNVNVGVRTAAALLDQFYEKRRLLI VSTPTARNLLYRLQLGMLQQAQCGLDLRHITVVELVGVFPTLIGRIGAKIMPPALALQLRLLLRIPL YSFSMVLVDKHGMDKERYVSLVMPVALFNLIDTFPLRKEEMVLQAEMSQTCNT		

	SEQ ID NO: 169	1798 bp	
NOV45b, CG171681-02 DNA Sequence	CTTGGTCTCTTCGGTCTCCTGCCGCCCCCGGGAAGCGCGCTGCGCTGCCGAGGCGAGCTAAGCGCCC GCTCGCC**ATG**GGGAGCCCCGCACATCGGCCCGCGCTGCTGCTGCTGCTGCCGCCTCTGCTGCTGCTG CTGCTGCTGCGCGTCCCGCCCAGCCGCAGCTTCCCAGATACCCCGTGGTGCTCCCCCATCAAGGTGA AGTATGGGGATGTGTACTGCAGGGCCCCTCAAGGAGGATACTACAAAACAGCCCTGGGAACCAGGTG CGACATTCGCTGCCAGAAGGGCTACGAGCTGCATGGCTCTTCCCTACTGATCTGCCAGTCAAACAAA CGATGGTCTGACAAGGTCATCTGCAAACAAAAGCGATGTCCTACCCTTGCCATGCCAGCAAATGGAG GGTTTAAGTGTGTAGATGGTGCCTACTTTAACTCCCGGTGTGAGTATTATTGTTCACCAGGATACAC GTTGAAAGGGGAGCGGACCGTCACATGTATGGACAACAAGGCCTGGAGCGGCCGGCCAGCCTCCTGT GTGGATATGGAACCTCCTAGAATCAAGTGCCCAAGTGTGAAGGAACGCATTGCAGAACCCAACAAAC TGACAGTCCGGGTGTCCTGGGAGACACCCGAAGGAAGAGACACAGCAGATGGAATTCTTACTGATGT CATTCTAAAAGGCCTCCCCCCAGGCTCCAACTTTCCAGAAGGAGACCACAAGATCCAGTACACAGTC TATGACAGAGCTGAGAATAAGGGCACTTGCAAATTTCGAGTTAAAGTAAGAGTCAAACGCTGTGGCA AACTCAATGCCCCAGAGAATGGTTACATGAAGTGCTCCAGCGACGGTGATAATTATGGAGCCACCTG TGAGTTCTCCTGCATCGGCGGCTATGAGCTCCAGGGTAGCCCTGCCCGAGTATGTCAATCCAACCTG GCTTGGTCTGGCACGGAGCCCACCTGTGCAGCCATGAACGTCAATGTGGGTGTCAGAACGGCAGCTG CACTTCTGGATCAGTTTTATGAGAAAAGGAGACTCCTCATTGTGTCCACACCCACAGCCCGAAACCT CCTTTACCGGCTCCAGCTAGGAATGCTGCAGCAAGCACAGTGTGGCCTTGATCTTCGACACATCACC GTGGTGGAGCTGGTGGGTGTGTTCCCGACTCTCATTGGCAGGATAGGAGCAAAGATTATGCCTCCAG CCCTAGCGCTGCAGCTCAGGCTGTTGCTGCGAATCCCACTCTACTCCTTCAGTATGGTGCTAGTGGA TAAGCATGGCATGGACAAAGAGCGCTATGTCTCCCTGGTGATGCCTGTGGCCCTGTTCAACCTGATT GACACTTTTCCCTTGAGAAAAGAAGAGATGGTCCTACAAGCCGAAATGAGCCAGACCTGTAACACCT GACATGATGGTTCCTCTCTTGGCAATTCCTCTTCATTGTCTACATAGTGACATGCACACGGGAAAGC CTTAAAAATATCCTTGATGTACAGATTTTATTTGTAATTTTAAAAGTCTATTTTATTATGAGCTTTC TTTGCACTTAAAAATTAGCATGCTGCTTTTTGTACTTGGAAGTGTTTCAAAAAATTATATGACCATA TTTACTCTTTCTAACTTTCTTTACTCCATCATGGCTGGTTGATTTGTAGAGAAATTAGAACCCATA ACCATACACAGGCTATCAACATGTTATTCAATGTGACACCTAACTCTTTTCTATTTGTTTTTTAAG TAAGACTTTTATTAATAAAACAAAATGTTTTGGAGCAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 75		ORF Stop: TGA at 1407

	SEQ ID NO: 170	444 aa	MW at 49381.1kD
NOV45b, CG171681-02 Protein Sequence	MGSPAHRPALLLLLPPLLLLLLLRVPPSRSFPDTPWCSPIKVKYGDVYCRAPQGGYYKTALGTRCDI RCQKGYELHGSSLLICQSNKRWSDKVICKQKRCPTLAMPANGGFKCVDGAYFNSRCEYYCSPGYTLK GERTVTCMDNKAWSGRPASCVDMEPPRIKCPSVKERIAEPNKLTVRVSWETPEGRDTADGILTDVIL KGLPPGSNFPEGDHKIQYTVYDRAENKGTCKFRVKVRVKRCGKLNAPENGYMKCSSDGDNYGATCEF SCIGGYELQGSPARVCQSNLAWSGTEPTCAAMNVNVGVRTAAALLDQFYEKRRLLIVSTPTARNLLY RLQLGMLQQAQCGLDLRHITVVELVGVFPTLIGRIGAKIMPPALALQLRLLLRIPLYSFSMVLVDKH GMDKERYVSLVMPVALFNLIDTFPLRKEEMVLQAEMSQTCNT		

	SEQ ID NO: 171	1795 bp	
NOV45c,	CTTGGTCTCTTCGGTCTCCTGCCGCCCCCGGGAAGCGCGCTGCGCTGCCGAGGCGAGCTAAGCGCCC		

CG171681-03 DNA Sequence	GCTCGCCATGGGGAGCCCCGCACATCGGCCCGCGCTGCTGCTGCTGCTGCCGCCTCTGCTGCTGCTG CTGCTGCGCGTCCCGCCCAGCCGCAGCTTCCCAGATACCCCGTGGTGCTCCCCCATCAAGGTGAAGT ATGGGGATGTGTACTGCAGGGCCCCTCAAGGAGGATACTACAAAACAGCCCTGGGAACCAGGTGCGA CATTCGCTGCCAGAAGGGCTACGAGCTGCATGGCTCTTCCCTACTGATCTGCCAGTCAAACAAACGA TGGTCTGACAAGGTCATCTGCAAACAAAAGCGATGTCCTACCCTTGCCATGCCAGCAAATGGAGGGT TTAAGTGTGTAGATGGTGCCTACTTTAACTCCCGGTGTGAGTATTATTGTTCACCAGGATACACGTT GAAAGGGGAGCGGACCGTCACATGTATGGACAACAAGGCCTGGAGCGGCCGGCCAGCCTCCTGTGTG GATATGGAACCTCCTAGAATCAAGTGCCCAAGTGTGAAGGAACGCATTGCAGAACCCAACAAACTGA CAGTCCGGGTGTCCTGGGAGACACCCGAAGGAAGAGACACAGCAGATGGAATTCTTACTGATGTCAT TCTAAAAGGCCTCCCCCCAGGCTCCAACTTTCCAGAAGGAGACCACAAGATCCAGTACACAGTCTAT GACAGAGCTGAGAATAAGGGCACTTGCAAATTTCGAGTTAAAGTAAGAGTCAAACGCTGTGGCAAAC TCAATGCCCCAGAGAATGGTTACATGAAGTGCTCCAGCGACGGTGATAATTATGGAGCCACCTGTGA GTTCTCCTGCATCGGCGGCTATGAGCTCCAGGGTAGCCCTGCCCGAGTATGTCAATCCAACCTGGCT TGGTCTGGCACGGAGCCCACCTGTGCAGCCATGAACGTCAATGTGGGTGTCAGAACGGCAGCTGCAC TTCTGGATCAGTTTTATGAGAAAAGGAGACTCCTCATTGTGTCCACACCCACAGCCCGAAACCTCCT TTACCGGCTCCAGCTAGGAATGCTGCAGCAAGCACAGTGTGGCCTTGATCTTCGACACATCACCGTG GTGGAGCTGGTGGGTGTGTTCCCGACTCTCATTGGCAGGATAGGAGCAAAGATTATGCCTCCAGCCC TAGCGCTGCAGCTCAGGCTGTTGCTGCGAATCCCACTCTACTCCTTCAGTATGGTGCTAGTGGATAA GCATGGCATGGACAAAGAGCGCTATGTCTCCCTGGTGATGCCTGTGGCCCTGTTCAACCTGATTGAC ACTTTTCCCTTGAGAAAAGAAGAGATGGTCCTACAAGCCGAAATGAGCCAGACCTGTAACACCTGAC ATGATGGTTCCTCTCTTGGCAATTCCTCTTCATTGTCTACATAGTGACATGCACACGGAAAGCCTT AAAAATATCCTTGATGTACAGATTTTATTTGTAATTTTAAAAGTCTATTTTATTATGAGCTTTCTTT GCACTTAAAAATTAGCATGCTGCTTTTTGTACTTGGAAGTGTTTCAAAAAATTATATGACCATATTT ACTCTTTCTAACTTTCTTTACTCCATCATGGCTGGTTGATTTTGTAGAGAAATTAGAACCCATAACC ATACACAGGCTATCAACATGTTATTCAATGTGACACCTAACTCTTTTCTATTTTGTTTTTAAGTAA GACTTTTATTAATAAAACAAAATGTTTTGGAGCAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 75		ORF Stop: TGA at 1404

	SEQ ID NO: 172	443 aa	MW at 49267.9kD
NOV45c, CG171681-03 Protein Sequence	MGSPAHRPALLLLLPPLLLLLLRVPPSRSFPDTPWCSPIKVKYGDVYCRAPQGGYYKTALGTRCDIR CQKGYELHGSSLLICQSNKRWSDKVICKQKRCPTLAMPANGGFKCVDGAYFNSRCEYYCSPGYTLKG ERTVTCMDNKAWSGRPASCVDMEPPRIKCPSVKERIAEPNKLTVRVSWETPEGRDTADGILTDVILK GLPPGSNFPEGDHKIQYTVYDRAENKGTCKFRVKVRVKRCGKLNAPENGYMKCSSDGDNYGATCEFS CIGGYELQGSPARVCQSNLAWSGTEPTCAAMNVNVGVRTAAALLDQFYEKRRLLIVSTPTARNLLYR LQLGMLQQAQCGLDLRHITVVELVGVFPTLIGRIGAKIMPPALALQLRLLLRIPLYSFSMVLVDKHG MDKERYVSLVMPVALFNLIDTFPLRKEEMVLQAEMSQTCNT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 45B.

Table 45B. Comparison of NOV45a against NOV45b and NOV45c.		
Protein Sequence	**NOV45a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV45b	33..321 156..444	273/289 (94%) 273/289 (94%)
NOV45c	33..321 155..443	273/289 (94%) 273/289 (94%)

Further analysis of the NOV45a protein yielded the following properties shown in Table 45C.

Table 45C. Protein Sequence Properties NOV45a

PSort analysis:	0.8200 probability located in outside; 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in lysosome (lumen)
SignalP analysis:	Cleavage site between residues 31 and 32

A search of the NOV45a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 45D.

Table 45D. Geneseq Results for NOV45a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV45a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB07747	A human cancer-associated protein-1 (CAP-1) - Homo sapiens, 465 aa. [WO200043508-A2, 27-JUL-2000]	33..319 178..464	148/287 (51%) 205/287 (70%)	7e-89
AAB59009	Breast and ovarian cancer associated antigen protein sequence SEQ ID 717 - Homo sapiens, 431 aa. [WO200055173-A1, 21-SEP-2000]	33..319 144..430	148/287 (51%) 205/287 (70%)	7e-89
ABB72149	Rat protein isolated from skin cells SEQ ID NO: 188 - Rattus sp, 118 aa. [WO200190357-A1, 29-NOV-2001]	88..203 3..118	71/116 (61%) 89/116 (76%)	3e-38
AAB55949	Skin cell protein, SEQ ID NO: 188 - Rattus sp, 118 aa. [WO200069884-A2, 23-NOV-2000]	88..203 3..118	71/116 (61%) 89/116 (76%)	3e-38
AAY76010	Rat DRS protein homologue, SEQ ID NO:188 - Rattus sp, 118 aa. [WO9955865-A1, 04-NOV-1999]	88..203 3..118	71/116 (61%) 89/116 (76%)	3e-38

In a BLAST search of public sequence datbases, the NOV45a protein was found to have homology to the proteins shown in the BLASTP data in Table 45E.

Table 45E. Public BLASTP Results for NOV45a				
Protein Accession Number	**Protein/Organism/Length**	**NOV45a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
P78539	Sushi repeat-containing protein SRPX precursor - Homo sapiens (Human), 464 aa.	33..321 176..464	289/289 (100%) 289/289 (100%)	e-168
Q63769	Sushi repeat-containing protein SRPX precursor (DRS protein) (Down-regulated by V-SRC) - Rattus norvegicus (Rat), 464 aa.	33..321 176..464	279/289 (96%) 286/289 (98%)	e-164
Q9R0M3	Sushi-repeat-containing protein - Mus musculus (Mouse), 464 aa.	33..320 176..463	276/288 (95%) 285/288 (98%)	e-163
Q9R0M2	Sushi-repeat-containing protein - Mus musculus (Mouse), 380 aa.	33..320 92..379	276/288 (95%) 285/288 (98%)	e-163
AAM73691	Sushi-repeat containing protein - Mus musculus (Mouse), 468 aa.	33..319 181..467	152/287 (52%) 203/287 (69%)	2e-89

PFam analysis predicts that the NOV45a protein contains the domains shown in the Table 45F.

Table 45F. Domain Analysis of NOV45a			
Pfam Domain	**NOV45a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
HYR	33..114	27/86 (31%) 78/86 (91%)	2.2e-34
sushi	119..174	19/64 (30%) 41/64 (64%)	2.7e-09

Example 46.

The NOV46 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 46A.

Table 46A. NOV46 Sequence Analysis			
	SEQ ID NO: 173	1785 bp	
NOV46a, CG173318-01 DNA Sequence	GTCGCCAGCTGAGGCGGTTTGTAAGTTTTGGGTCGCAGTATGCTAGAATTTTGAGGCTCCCTTCTGA TGAAAATTGAGCTGTCCATGCAGCCATGGAACCCGGGTTACAGCAGTGAGGGGGCCACGGCTCAAGA AACTTACACATGTCCAAAAATGATTGAGATGGAGCAGGCGGAGGCCCAGCTTGCTGAGTTAGACCTG CTAGCCAGTATGTTCCCTGGTGAGAATGAGCTCATAGTGAATGACCAGCTGGCTGTAGCAGAACTGA AAGATTGTATTGAAAAGAAGACAATGGAGGGGCGATCTTCAAAAGTCTACTTTACTATCAATATGAA CCTGGATGTATCTGACGAAAAAATGGTAATTCAGTTTTGCTTTTAGAGGGATTGAAACATGTTGAGA CTTAAAACATTGGTTAGTGCACTTTTTCTTCTTCTCTTTAATCAGGCGATGTTTTCTCTGGCCTGTA TTCTTCCCTTTAAATACCCGGCAGTTCTGCCTGAAATTACTGTCAGATCAGTATTATTGAGTAGATC CCAGCAGACTCAGCTGAACACAGATCTGACTGCATTCCTGCAAAAACATTGTCATGGAGATGTTTGT ATACTGAATGCCACAGAGTGGGTTAGAGAACACGCCTCTGGCTATGTCAGCAGAGATACTTCATCTT CACCCACCACAGGAAGCACAGTCCAGTCAGTTGACCTCATCTTCACGAGACTCTGGATCTACAGCCA TCATATCTATAACAAATGCAAAAGAAAGAATATTCTAGAGTGGGCAAAGGAGCTTTCCCTGTCTGGG TTTAGCATGCCTGGAAAACCTGGTGTTGTTTGTGTGGAAGGCCCACAAAGTGCCTGTGAAGAATTCT GGTCAAGACTCAGAAAATTAAACTGGAAGAGAATTTTAATTCGCCATCGAGAAGACATTCCTTTTGA TGGTACAAATGATGAAACGGAAAGACAAAGGAAATTTTCCATTTTTGAAGAAAAAGTGTTCAGTGTT AATGGAGCCAGGGGAAACCACATGGACTTTGGTCAGCTCTATCAGTTCTTAAACACCAAAGGATGTG GGGATGTTTTCCAGATGTTCTTTGGTGTAGAAGGACAATGACATCAAGAGTAGTTGAAAGTATCTTG CCACTGTTGGCCTTTTGATTTTTTTTTCCCACTTTTTCTTGAAAGATTAAGTAATTTTATTTTAGTT CCATTCTAGAATGTTGGGGAGTGGGCACAAGAAAAAATAGTATAGCTGAAATGCATCTGTTAAAAA TGTCATGATTGAAAGCAGAACTGAGTTTCAAATTACAACCTTAAAATTGTTGTTAGATATTTCTTCA CATATCAGCTGCCCATTTTGAAAAAGAAATTATCCATAAAGGTAATGTTGGTGCTCCAATTTGCCAG CCATTCCCAACCCCCTTCTCCCTTACCTGCCTTCACTAAAGAACCCAGAAAAGCTAATTGCTCCCCT TTCAGCCTCTGTTGCAACTAACAACTCTCAGTGGCCTCAGGACACAGCTTTGGCCTTGGGAATTCTG GGAAAACTTTTACTTCCTGATTAAAGATACATATGCAGCTAGGCCACCTCCTCCCCCCCTTACTGCC ATAAACACCAAAGTGATGACTGGAGCTGGAGGAGTTATTTGAACCACGACGGAAGGGCCAAGAGAAC CACGAAGATGCCAGTTGCCACATTGTTGAGCTGCTGACCCAACACCAGCCATTGCCTGTCTCTAAAC ATCTTATGAAATAAAACCAATTTTGTTTAAAAAAAAAAAAAAA		
	ORF Start: ATG at 394		ORF Stop: TGA at 1111

	SEQ ID NO: 174	239 aa	MW at 27409.3kD
NOV46a, CG173318-01 Protein Sequence	MLRLKTLVSALFLLLFNQAMFSLACILPFKYPAVLPEITVRSVLLSRSQQTQLNTDLTAFLQKHCHG DVCILNATEWVREHASGYVSRDTSSSPTTGSTVQSVDLIFTRLWIYSHHIYNKCKRKNILEWAKELS LSGFSMPGKPGVVCVEGPQSACEEFWSRLRKLNWKRILIRHREDIPFDGTNDETERQRKFSIFEEKV FSVNGARGNHMDFGQLYQFLNTKGCGDVFQMFFGVEGQ		

Further analysis of the NOV46a protein yielded the following properties shown in Table 46B.

Table 46B. Protein Sequence Properties NOV46a	
PSort analysis:	0.8000 probability located in outside; 0.2726 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 23 and 24

A search of the NOV46a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 46C.

Table 46C. Geneseq Results for NOV46a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE15253	Human RNA metabolism protein-16 (RMEP-16) - Homo sapiens, 319 aa. [WO200183524-A2, 08-NOV-2001]	19..239 99..319	221/221 (100%) 221/221 (100%)	e-131
AAM78405	Human protein SEQ ID NO 1067 - Homo sapiens, 319 aa. [WO200157190-A2, 09-AUG-2001]	19..239 99..319	221/221 (100%) 221/221 (100%)	e-131
AAM79389	Human protein SEQ ID NO 3035 - Homo sapiens, 354 aa. [WO200157190-A2, 09-AUG-2001]	19..236 137..354	215/218 (98%) 216/218 (98%)	e-127
ABB11888	Human novel protein, SEQ ID NO:2258 - Homo sapiens, 354 aa. [WO200157188-A2, 09-AUG-2001]	19..236 137..354	215/218 (98%) 216/218 (98%)	e-127
AAB58229	Lung cancer associated polypeptide sequence SEQ ID 567 - Homo sapiens, 305 aa. [WO200055180-A2, 21-SEP-2000]	19..167 103..251	147/149 (98%) 147/149 (98%)	9e-84

In a BLAST search of public sequence datbases, the NOV46a protein was found to have homology to the proteins shown in the BLASTP data in Table 46D.

Table 46D. Public BLASTP Results for NOV46a

Protein Accession Number	Protein/Organism/Length	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P57060	Protein C21orf6 (GL011) - Homo sapiens (Human), 319 aa.	19..239 99..319	221/221 (100%) 221/221 (100%)	e-130
Q99M03	Similar to open reading frame 5 - Mus musculus (Mouse), 290 aa.	21..239 72..290	182/219 (83%) 192/219 (87%)	e-105
Q9DCJ3	Open reading frame 5 - Mus musculus (Mouse), 244 aa.	21..239 26..244	182/219 (83%) 192/219 (87%)	e-105
Q9JLH4	Orf5 protein - Mus musculus (Mouse), 291 aa.	21..239 73..291	181/219 (82%) 192/219 (87%)	e-105

Q9D9S3	1700030C20Rik protein - Mus musculus (Mouse), 292 aa.	23..239 72..288	85/222 (38%) 127/222 (56%)	3e-38

PFam analysis predicts that the NOV46a protein contains the domains shown in the Table 46E.

Table 46E. Domain Analysis of NOV46a			
Pfam Domain	**NOV46a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**

Example 47.

The NOV47 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 47A.

Table 47A. NOV47 Sequence Analysis			
	SEQ ID NO: 175	6373 bp	
NOV47a, CG51595-01 DNA Sequence	GACAGAGTGCAGCCTTTTCAGACTCTGTGACACAGTTCCCCTTTTGCAAAAATACTTAGCGAGGATC ATTACTTTCCAACAGTCGTGTCCAGAGACCTACTTTGTAACACCGCAGGGAAGTTAATGTACTAGGT CTTGAAAGGTCTTTCTGGAATGTGCAGTAACTTGTAGTTTTCTTCTAGTAGCACTGCTAATTTTTGT GTTATAATTTTTGTAGGTCCATGGGGCCGATGT**ATG**GGAGATGAATGTGGTCCCGGAGGCATCCAAA CGAGGGCTGTGTGGTGTGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGA GAGACCCAATAACCAGCAGAATTGTTTCAAAGTTTGCGATTGGCACAAAGAGTTGTACGACTGGAGA CTGGGACCTTGGAATCAGTGTCAGCCCGTGATTTCAAAAAGCCTAGAGAAACCTCTTGAGTGCATTA AGGGGGAAGAAGGTATTCAGGTGAGGGAGATAGCGTGCATCCAGAAAGACAAAGACATTCCTGCGGA GGATATCATCTGTGAGTACTTTGAGCCCAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTTGCCAG CAAGATTGCATCGTGTCTGAATTTTCTGCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGGGCTCC AGCACCGGACGCGTCATGTGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGA GTTCCAGGTGTGCCAATCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCC TGGAGCACCTGCTCAATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAG AACGGGAAAAGGACCGCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGAG AAACAGAAACAGGCAGAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATCAGACCAGA GAGGTTATGTGCATTAACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAAGCTTC CAATGACCTTCCAGTCCTGTGTGATCACCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGTGGAGCCC CTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACACGAACCATCAGG CAGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAG GAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGGA CCCTTTGCTCAGTCAGCAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAG ACCCGAGAGGTGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACA AAGAAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGCTGTG CCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTTATGAA AACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAATGAGCCCA CTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAGCCTGC CTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGGAGTGTGGTCCA GGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAGACAGCTGTGCA GAGATGCCATCTTCCCCATCCCTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTCAGCAC ATGGTCTACGTGGTCCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGA GCACGATCCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGC AAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCTGGGGCCA GTGCATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATGGGGAGGCCTCCTGC TCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGACCCAAAA AATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCTTGTAAGAAGGACTGTAT TGTGACCCCATATAGTGACTGGACATCATGCCCCTCTTCGTGTAAAGAAGGGGACTCCAGTATCAGG		

	AAGCAGTCTAGGCATCGGGTCATCATTCAGCTGCCAGCCAACGGGGGCCGAGACTGCACAGATCCCC TCTATGAAGAGAAGGCCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATG GCGCAGATGCCAATTAGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGG CCTGGGCGACAGGCAAGAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATGAGT GCCTACAGTATGCAGGCCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCCAGGATGACTG TCAATTGACCAGCTGGTCCAAGTTTTCTTCATGCAATGGAGACTGTGGTGCAGTTAGGACCAGAAAG CGCACTCTTGTTGGAAAAAGTAAAAAGAAGGAAAAATGTAAAAATTCCCATTTGTATCCCCTGATTG AGACTCAGTATTGTCCTTGTGACAAATATAATGCACAACCTGTGGGGAACTGGTCAGACTGTATTTT ACCAGAGGGAAAAGTGGAAGTGTTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAA GGATATCGTTACCAAGCAATGGCATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTA ACAGCCATGGTTACATTGAGGAGGCCTGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTG GTCCAACTGGTCGCGCTGCAGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGT GAAAAACCATATAATGGAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATG AGGTTGTCCCATGCCACAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAA GGTGACCTTTGTGAATATGCGGGAGAACTGTGGAGAGGGCGTGCAAACCCGAAAAGTGAGATGCATG CAGAATACAGCAGATGGCCCTTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAAGAGATGCCCC TGGGCTCTAGAGTGTGCAAATTACCATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGAC CCAATGTGTTTTGCCTTGCAATCAAAGCAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCA GCTGATGAAGGAAGATCTTGCCCTAATGCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCT ACCACTATGATTATAATGTAACAGACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAA TGGAATAAAAACAAGGATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGT GAAGCGCTTGGCTTGGAGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACT GTCAGCTTTCTGATTGGTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCCTCACAGGAAAAATGAT CCGAAGACGAACAGTGACCCAGCCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCAG TCCAAACCCTGCCCAGTGAAGCCTTGTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGC AGGAGGCCCAGTGTGGAGAAGGGACCAGAACAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGC TGATGATTTCAGCAAAGTGGTGGATGAGGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAAT AAAAATATGGTTCTGGAGGAATCCTGCAGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGT CTTCCTGGAGCCTGTGTCAGCTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGT CAGATCCAGACCGGTGATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAA ACAAAATCATGTTATGATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTT CCCGAACAGTGTGGTGTCAAAGGTCAGATGGTATAAATGTAACAGGGGGCTGCTTGGTGATGAGCCA GCCTGATGCCGACAGGTCTTGTAACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAA ACATGCCATTGTGAAGAAGGGTACACTGAAGTCATGTCTTCTAACAGCACCCTTGAGCAATGCACAC TTATCCCCGTGGTGGTATTACCCACCATGGAGGACAAAAGAGGAGATGTGAAAACCAGTCGGGCTGT ACATCCAACCCAACCCTCCAGTAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGG CCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTTGTGTTACTCATCTTTA TTGTCTCCATGATTTATCTAGCTTGCAAAAAGCCAAAGAAACCCCAAAGAAGGCAAAACAACCGACT GAAACCTTTAACCTTAGCCTATGATGGAGATGCCGACATGTAACATATAACTTTTCCTGGCAACAAC CAGTTTCGGCTTTCTGACTTCATAGATGTCCAGAGGCCACAACAAATGTATCCAAACTGTGTGGATT AAAATATATTTTAATTTTTAAAAATGGCATCATAAAGACAAGAGTGAAAATCATACTGCCACTGGAG ATATTTAAGACAGTACCACTTATATACAGACCATCAACCGTGAGAATTATAGGAGATTTAGCTGAAT ACATGCTGCATTCTGAAAGTTTTATGTCATCTTTTCTGAAATCTACCGACTGAAAAACCACTTTCAT CTCTAAAAAATAATGGTGGAATTGGCCAGTTAGGATGCCTGATACAAGACCGTCTGCAGTGTTAATC CATAAAACTTCCTAGCATGAAGAGTTTCTACCAAGATCTCCACAATACTATGGTCAAATTAACATGT GTACTCAGTTGAATGACACACATTATGTCAGATTATGTACTTGCTAATAAGCAATTTTAACAATGCA TAACAAATAAACTCTAAGCTAAGCAGAAAATCCACTGAATAAATTCAGCATCTTGGTGGTCGATGGT AGATTTTATTGACCTGCATTTCAGAGACAAAGCCTCTTTTTTAAGACTTCTTGTCTCTCTCCAAAGT AAGAATGCTGGACAAGTACTAGTGTCTTAGAAGAACGAGTCCTCAAGTTCAGTATTTTATAGTGGTA ATTGTCTGGAAAACTAATTTACTTGTGTTAATACAATACGTTTCTACTTTCCCTGATTTTCAAACTG GTTGCCTGCATCTTTTTTGCTATATGGAAGGCACATTTTTGCACTATATTAGTGCAGCACGATAGGC GCTTAACCAGTATTGCCATAGAAACTGCCTCTTTTCATGTGGGATGAAGACATCTGTGCCAAGAGTG GCATGAAGACATTTGCAAGTTCTTGTATCCTGAAGAGAGTAAAGTTCAGTTTGGATGGCAGCAAGAT GAAATCAGCTATTACACCTGCTGTACACACACTTCCTCATCACTGCAGCCATTGTGAAATTGACAAC ATGGCGGTAATTTAAGTGTTGAAGTCCCTAACCCCTTAACCCTCTAAAAGGTGGATTCCTCTAGTTG GTTTGTAATTGTTCTTTGAAGGCTGTTTATGACTAGATTTTTATATTTGTTATCTTTGTTAAGAAAA AAAAAAGAAAAAGGAACTGGATGTCTTTTTAATTTTGAGCAGATGGAGAAAATAAATAATGTATCAA TGACCTTTGTAACTAAAGGAAAAAAAAAAAAAAATGTGGATTTTCCTTTCTCTCTGATTTCCCAGTTT CAGATTGAATGTCTGTCTTGCAGGCAGTTATTTCAAAATCCATAGTCTTTNGCCTTTCTCACTGGCA AAATTTGA		
	ORF Start: ATG at 235		ORF Stop: TAA at 4999

	SEQ ID NO: 176	1588 aa	MW at 178042.1kD
NOV47a, CG51595-01 Protein Sequence	MGDECGPGGIQTRAVWCAHVEGWTTLHTNCKQAERPNNQQNCFKVCDWHKELYDWRLGPWNQCQPVI SKSLEKPLECIKGEEGIQVREIACIQKDKDIPAEDIICEYFEPKPLLEQACLIPCQQDCIVSEFSAW SECSKTCGSGLQHRTRHVVAPPQFGGSGCPNLTEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHSRQ VRQARRRGKNKEREKDRSKGVKDPEARELIKKKRNRNRQNRQENKYWDIQIGYQTREVMCINKTGKA ADLSFCQQEKLPMTFQSCVITKECQVSEWSEWSPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKECP EFEEKEPCLSQGDGVVPCATYGWRTTEWTECRVDPLLSQQDKRRGNQTALCGGGIQTREVYCVQANE NLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQLCHIPCPTECEVSPWSAWGPCTYENCNDQQGKKGF KLRKRRITNEPTGGSGVTGNCPHLLEAIPCEEPACYDWKAVRLGDCEPDNGKECGPGTQVQEVVCIN SDGEEVDROLCRDAIFPIPVACDAPCPKDCVLSTWSTWSSCSHTCSGKTTEGKOIRARSILAYAGEE		

	GGIRCPNSSALQEVRSCNEHPCTVYHWQTGPWGQCIEDTSVSSFNTTTTWNGEASCSVGMQTRKVIC VRVNVGQVGPKKCPESLRPETVRPCLLPCKKDCIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVIIQL PANGGRDCTDPLYEEKACEAPQACQSYRWKTHKWRRCQLVPWSVQQDSPGAQEGCGPGRQARAITCR KQDGGQAGIHECLQYAGPVPALTQACQIPCQDDCQLTSWSKFSSCNGDCGAVRTRKRTLVGKSKKKE KCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLLGMKVQGDIKECGQGYRYQAMACYD QNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKSCGSGVKVRSKWLREKPYNGGRPCP KLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMRENCGEGVQTRKVRCMQNTADGPSEHV EDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQSSFRQRSADPIRQPADEGRSCPNAV EKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDCVRSDGKSVDLKYCEALGLEKNWQM NTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPFQGDGRPCPSLMDQSKPCPVKPCYR WQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDEEFCADIELIIDGNKNMVLEESCSQ PCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQELENQHLCPEQMLETKSCYDGQCYE YKWMASAWKGSSRTVWCQRSDGINVTGGCLVMSQPDADRSCNPPCSQPHSYCSETKTCHCEEGYTEV MSSNSTLEQCTLIPVVVLPTMEDKRGDVKTSRAVHPTQPSSNPAGRGRTWFLQPFGPDGRLKTWVYG VAAGAFVLLIFIVSMIYLACKKPKKPQRRQNNRLKPLTLAYDGDADM

	SEQ ID NO: 177	1401 bp
NOV47b, CG51595-03 DNA Sequence	GAGTGGAGCCCCTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACAC GAACCATCAGGCAGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTG TTTGTCTCAAGGAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAG TGCCGTGTGGACCCTTTGCTCAGTCAGCAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAG GGGGCATCCAGACCCGAGAGGTGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTAC CCACAAGAACAAAGAAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACT ACACAGCTGTGCCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTT GTACTTATGAAAACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTAC CAATGAGCCCACTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGT GAAGAGCCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGG AGTGTGGTCCAGGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAG ACAGCTGTGCAGAGATGCCATCTTCCCCATCCCTGTGGCCTGTGATGCCCCGTCCCCGAAAGACTGT GTGCTCAGCACATGGTCTACGTGGTCCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGA AACAGATACGAGCACGATCCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAG CAGTGCTTTGCAAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGT CCCTGGGGCCAGTGCATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATGGGG AGGCCTCCTGCTCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCCAAGT GGGACCCAAAAAATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCTTGTAAG AAGGAGTGTATTGTGACCCCATATAGTGACTGGACATCATGCCCCTCTTCGTGTAAAGAAGGGGACT CCAGTATCAGGAAGCAGTCTAGGCATCGGGTCATCATTCAGCTGCCAGCCAACGGGGGCCGAGACTG CACAGATCCCCTCTATGAAGAGAAGGCCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGG	
	ORF Start: at 1	ORF Stop: end of sequence

	SEQ ID NO: 178	467 aa	MW at 51476.5kD
NOV47b, CG51595-03 Protein Sequence	EWSPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKECPEFEEKEPCLSQGDGVVPCATYGWRTTEWTE CRVDPLLSQQDKRRGNQTALCGGGIQTREVYCVQANENLLSQLSTHKNKEASKPMDLKLCTGPIPNT TQLCHIPCPTECEVSPWSAWGPCTYENCNDQQGKKGFKLRKRRITNEPTGGSGVTGNCPHLLEAIPC EEPACYDWKAVRLGDCEPDNGKECGPGTQVQEVVCINSDGEEVDRQLCRDAIFPIPVACDAPSPKDC VLSTWSTWSSCSHTCSGKTTEGKQIRARSILAYAGEEGGIRCPNSSALQEVRSCNEHPCTVYHWQTG PWGQCIEDTSVSSFNTTTTWNGEASCSVGMQTRKVICVRVNVGQVGPKKCPESLRPETVRPCLLPCK KECIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVIIQLPANGGRDCTDPLYEEKACEAPQACQSYR		

	SEQ ID NO: 179	1713 bp
NOV47c, CG51595-04 DNA Sequence	TGCAATGGAGACTGTGGTGCAGTTAGGACCAGAAAGCGCACTCTTGTTGGAAAAAGTAAAAAGAAGG AAAAATGTAAAAATTCCCATTTGTATCCCCTGATTGAGACTCAGTATTGTCCTTGTGACAAATATAA TGCACAACCTGTGGGGAACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTGGAAGTGTTGCTGGGA ATGAAAGTACAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAGCAATGGCATGCTACG ATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACATTGAGGAGGCCTGCAT CATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACTGGTCGCGCTGCAGCAAGTCCTGT GGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTGAAAAACCATATAATGGAGGAAGGCCTTGCC CCAAACTGGACCATGTCAACCAGGCACAGGTGTATGAGGTTGTCCCATGCCACAGTGACTGCAACCA GTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAGAACTGT GGAGAGGGCGTGCAAACCCGAAAAGTGAGATGCATGCAGAATACAGCAGATGGCCCTTCTGAACATG TAGAGGATTACCTCTGTGACCCAGAAGAGATGCCCCTGGGCTCTAGAGTGTGCAAATTACCATGCCC TGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCTTGCAATCAAAGCAGT TTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGATCTTGCCCTAATGCTG TTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATTATAATGTAACAGACTGGAG TACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAATAAAAACAAGGATGTTGGATTGTGTT	

	CGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAGAACTGGCAGA TGAACACGTCCTGCATGGTGGAATGCCCTGTGAACTGTCAGCTTTCTGATTGGTCTCCTTGGTCAGA ATGTTCTCAAACATGTGGCCTCACAGGAAAAATGATCCGAAGACGAACAGTGACCCAGCCCTTTCAA GGTGATGGAAGACCATGCCCTTCCCTGATGGACCAGTCCAAACCCTGCCCAGTGAAGCCTTGTTATC GGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGGAGAAGGGACCAGAAC AAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAAGTGGTGGATGAGGAA TTCTGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGGAGGAATCCTGCAGCC AGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGAGCCTGTGTCAGCTGACCTGCGT GAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCAGATCCAGACCGGTGATTATACAAGAACTA GAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAATCATGTTATGATGGACAGTGCTATG AATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTTCC	
	ORF Start: at 1	ORF Stop: end of sequence

	SEQ ID NO: 180	571 aa	MW at 64468.4kD
NOV47c, CG51595-04 Protein Sequence	CNGDCGAVRTRKRTLVGKSKKKEKCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLLG MKVQGDIKECGQGYRYQAMACYDQNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKSC GSGVKVRSKWLREKPYNGGRPCPKLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMRENC GEGVQTRKVRCMQNTADGPSEHVEDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQSS FRQRSADPIRQPADEGRSCPNAVEKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDCV RSDGKSVDLKYCEALGLEKNWQMNTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPFQ GDGRPCPSLMDQSKPCPVKPCYRWQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDEE FCADIELIIDGNKNMVLEESCSQPCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQEL ENQHLCPEQMLETKSCYDGQCYEYKWMASAWKGSS		

	SEQ ID NO: 181	4881 bp	
NOV47d, CG51595-06 DNA Sequence	CGTCCATGGGGCCGATGTATGGGAGATGAATGTGGTCCCGGAGGCATCCAAACGAGGGCTGTGTGGT GTGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGAGAGACCCAATAACCA GCAGAATTGTTTCAAAGTTTGCGATTGGCACAAAGAGTTGTACGACTGGAGACTGGGACCTTGGAAT CAGTGTCAGCCCGTGATTTCAAAAAGCCTAGAGAAACCTCTTGAGTGCATTAAGGGGGAAGAAGGTA TTCAGGTGAGGGAGATAGCGTGCATCCAGAAAGACAAAGACATTCCTGCGGAGGATATCATCTGTGA GTACTTTGAGCCCAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTTGCCAGCAAGATTGCATCGTG TCTGAATTTTCTGCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGGGCTCCAGCACCGGACGCGTC ATGTGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAGGTGTGCCA ATCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCCTGGAGCACCTGCTCA ATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAGAACGGGAAAAGGACC GCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGAGAAACAGAAACAGGCA GAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATCAGACCAGAGAGGTTATGTGCATT AACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAAGCTTCCAATGACCTTCCAGT CCTGTGTGATCACCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGTGGAGCCCCTGCTCAAAAACATG CCATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACACGAACCATCAGGCAGTTTCCCATTGGC AGTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAGGAGATGGAGTTGTCC CCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGGACCCTTTGCTCAGTCA GCAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAGACCCGAGAGGTGTAC TGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACAAAGAAGCCTCAAAGC CAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGCTGTGCCACATTCCTTGTCC AACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTTATGAAAACTGTAATGATCAG CAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAATGAGCCCACTGGAGGCTCTGGGG TAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAGCCTGCCTGTTATGACTGGAA AGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGGAGTGTGGTCCAGGCACGCAAGTTCAA GAGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCC CCATCCCTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTCAGCACATGGTCTACGTGGTC CTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGAGCACGATCCATTCTG GCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGCAAGAAGTACGAAGCT GTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCTGGGGCCAGTGCATTGAGGACAC CTCAGTATCGTCCTTCAACACAACTACGACTTGGAATGGGGAGGCCTCCTGCTCTGTCGGCATGCAG ACAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAAAATGTCCTGAAAGCC TTCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCTTGTAAGAAGGACTGTATTGTGACCCCATATAG TGACTGGACATCATGCCCCTCTTCGTGTAAAGAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCAT CGGGTCATCATTCAGCTGCCAGCCAACGGGGGCCGAGACTGCACAGATCCCCTCTATGAAGAGAAGG CCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATGGCGCAGATGCCAATT AGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGGCCTGGGCGACAGGCA AGAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGCAG GCCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCCAGGATGACTGTCAATTGACCAGCTG GTCCAAGTTTTCTTCATGCAATGGAGACTGTGGTGCAGTTAGGACCAGAAAGCGCACTCTTGTTGGA AAAAGTAAAAAGAAGGAAAAATGTAAAAATTCCCATTTGTATCCCCTGATTGAGACTCAGTATTGTC CTTGTGACAAATATAATGCACAACCTGTGGGGAACTGGTCAGACTGTATTTTACCAGAGGGAAAAGT GGAAGTGTTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAA GCAATGGCATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACA TTGAGGAGGCCTGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACTGGTCGCG		

	CTGCAGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTGAAAAACCATATAAT GGAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATGAGGTTGTCCCATGCC ACAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAA TATGCGGGAGAACTGTGGAGAGGGCGTGCAAACCCGAAAAGTGAGATGCATGCAGAATACAGCAGAT GGCCCTTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAAGAGATGCCCCTGGGCTCTAGAGTGT GCAAATTACCATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCC TTGCAATCAAAGCAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGA TCTTGCCCTAATGCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATTATA ATGTAACAGACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAATAAAAACAAG GATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTG GAGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACTGTCAGCTTTCTGATT GGTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCCTCACAGGAAAAATGATCCGAAGACGAACAGT GACCCAGCCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCAGTCCAAACCCTGCCCA GTGAAGCCTTGTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTG GAGAAGGGACCAGAACAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAA AGTGGTGGATGAGGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTG GAGGAATCCTGCAGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGAGCCTGT GTCAGCTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCAGATCCAGACCGGT GATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAATCATGTTAT GATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTTCCCGAACAGTGTGGT GTCAAAGGTCAGATGGTATAAATGTAACAGGGGGCTGCTTGGTGATGAGCCAGCCTGATGCCGACAG GTCTTGTAACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAAACATGCCATTGTGAA GAAGGGTACACTGAAGTCATGTCTTCTAACAGCACCCTTGAGCAATGCACACTTATCCCCGTGGTGG TATTACCCACCATGGAGGACAAAAGAGGAGATGTGAAAACCAGTCGGGCTGTACATCCAACCCAACC CTCCAGTAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGGCCAGCAAAAAGCCAA AGAAACCCCAAAGAAGGCAAAACAACCGAC**TGA**AACCTTTAACCTTAGCCTATGATGGAGATGCCGA CATGTAACATATAACTTTTCCTGGCAACAACCAGTTTCGGCTTTCTGACTTCATAGATGTCCAGAGG CCACAACAAATGTATCCAAACTGTGTGGATTAAAATATATTTTAATTTTTAAAAATGGCATCATAAA GACAAGAGTGAAAATCATACTGCCACTGGAGATATTTAAGACAGTACCACTTATATA		
	ORF Start: ATG at 19		ORF Stop: TGA at 4654

	SEQ ID NO: 182	1545 aa	MW at 173146.2kD
NOV47d, CG51595-06 Protein Sequence	MGDECGPGGIQTRAVWCAHVEGWTTLHTNCKQAERPNNQQNCFKVCDWHKELYDWRLGPWNQCQPVI SKSLEKPLECIKGEEGIQVREIACIQKDKDIPAEDIICEYFEPKPLLEQACLIPCQQDCIVSEFSAW SECSKTCGSGLQHRTRHVVAPPQFGGSGCPNLTEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHSRQ VRQARRRGKNKEREKDRSKGVKDPEARELIKKKRNRNRQNRQENKYWDIQIGYQTREVMCINKTGKA ADLSFCQQEKLPMTFQSCVITKECQVSEWSEWSPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKECP EFEEKEPCLSQGDGVVPCATYGWRTTEWTECRVDPLLSQQDKRRGNQTALCGGGIQTREVYCVQANE NLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQLCHIPCPTECEVSPWSAWGPCTYENCNDQQGKKGF KLRKRRITNEPTGGSGVTGNCPHLLEAIPCEEPACYDWKAVRLGDCEPDNGKECGPGTQVQEVVCIN SDGEEVDRQLCRDAIFPIPVACDAPCPKDCVLSTWSTWSSCSHTCSGKTTEGKQIRARSILAYAGEE GGIRCPNSSALQEVRSCNEHPCTVYHWQTGPWGQCIEDTSVSSFNTTTTWNGEASCSVGMQTRKVIC VRVNVGQVGPKKCPESLRPETVRPCLLPCKKDCIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVIIQL PANGGRDCTDPLYEEKACEAPQACQSYRWKTHKWRRCQLVPWSVQQDSPGAQEGCGPGRQARAITCR KQDGGQAGIHECLQYAGPVPALTQACQIPCQDDCQLTSWSKFSSCNGDCGAVRTRKRTLVGKSKKKE KCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLLGMKVQGDIKECGQGYRYQAMACYD QNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKSCGSGVKVRSKWLREKPYNGGRPCP KLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMRENCGEGVQTRKVRCMQNTADGPSEHV EDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQSSFRQRSADPIRQPADEGRSCPNAV EKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDCVRSDGKSVDLKYCEALGLEKNWQM NTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPFQGDGRPCPSLMDQSKPCPVKPCYR WQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDEEFCADIELIIDGNKNMVLEESCSQ PCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQELENQHLCPEQMLETKSCYDGQCYE YKWMASAWKGSSRTVWCQRSDGINVTGGCLVMSQPDADRSCNPPCSQPHSYCSETKTCHCEEGYTEV MSSNSTLEQCTLIPVVVLPTMEDKRGDVKTSRAVHPTQPSSNPAGRGRTWFLQPFGPAKSQRNPKEG KTTD		

	SEQ ID NO: 183	4679 bp	
NOV47e, CG51595-07 DNA Sequence	GTCCATGGGGCCGATGTATGGGAGATGAATGTGGTCCCGGAGGCATCCAAACGAGGGCTGTGTGGTG TGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGAGAGACCCAATAACCAG CAGAATTGTTTCAAAGTTTGCGATTGGCACAAAGAGTTGTACGACTGGAGACTGGGACCTTGGAATC AGTGTCAGCCCGTGATTTCAAAAAGCCTAGAGAAACCTCTTGAGTGCATTAAGGGGGAAGAAGGTAT TCAGGTGAGGGAGATAGCGTGCATCCAGAAAGACAAAGACATTCCTGCGGAGGATATCATCTGTGAG TACTTTGAGCCCAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTTGCCAGCAAGATTGCATCGTGT CTGAATTTTCTGCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGGGCTCCAGCACCGGACGCGTCA TGTGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAGGTGTGCCAA TCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCCTGGAGCACCTGCTCAA TGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAGAACGGAAAAAGGACCG CAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGAGAAACAGAAACAGGCAG		

	AACAGACAAGAGAACAAATATTGGACATCCAGATTGGATATCAGACCAGAGAGGTTATGTGCATTA ACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAAGCTTCCAATGACCTTCCAGTC CTGTGTGATCACCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGTGGAGCCCCTGCTCAAAAACATGC CATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACACGAACCATCAGGCAGTTTCCCATTGGCA GTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAGGAGATGGAGTTGTCCC CTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGGACCCTTTGCTCAGTCAG CAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAGACCCGAGAGGTGTACT GCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACAAAGAAGCCTCAAAGCC AATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGCTGTGCCACATTCCTTGTCCA ACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTTATGAAAACTGTAATGATCAGC AAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAATGAGCCCACTGGAGGCTCTGGGGT AACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAGCCTGCCTGTTATGACTGGAAA GCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGGAGTGTGGTCCAGGCACGCAAGTTCAAG AGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCCC CATCCCTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTCAGCACATGGTCTACGTGGTCC TCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGAGCACGATCCATTCTGG CCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGCAAGAAGTACGAAGCTG TAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCTGGGGCCAGTGCATTGAGGACACC TCAGTATCGTCCTTCAACACAACTACGACTTGGAATGGGGAGGCCTCCTGCTCTGTCGGCATGCAGA CAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGACCCAAAAAATGTCCTGAAAGCCT TCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCTTGTAAGAAGGACTGTATTGTGACCCCATATAGT GACTGGACATCATGCCCCTCTTCGTGTAAAGAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCATC GGGTCATCATTCAGCTGCCAGCCAACGGGGGCCGAGACTGCACAGATCCCCTCTATGAAGAGAAGGC CTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATGGCGCAGATGCCAATTA GTCCCTTGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGGCCTGGGCGACAGGCAA GAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGCAGG CCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCCAGGATGACTGTCAATTGACCAGCTGG TCCAAGTTTTCTTCATGCAATGGAGACTGTGGTGCAGTTAGGACCAGAAAGCGCACTCTTGTTGGAA AAAGTAAAAAGAAGGAAAAATGTAAAAATTCCCATTTGTATCCCCTGATTGAGACTCAGTATTGTCC TTGTGACAAATATAATGCACAACCTGTGGGAACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTG GAAGTGTTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAG CAATGGCATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACAT TGAGGAGGCCTGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACTGGTCGCGC TGCAGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTGAAAAACCATATAATG GAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATGAGGTTGTCCCATGCCA CAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAAT ATGCGGGAGAACTGTGGAGAGGGCGTGCAAACCCGAAAAGTGAGATGCATGCAGAATACAGCAGATG GCCCTTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAAGAGATGCCCCTGGGCTCTAGAGTGTG CAAATTACCATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCT TGCAATCAAAGCAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGAT CTTGCCCTAATGCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATTATAA TGTAACAGACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAATAAAAACAAGG ATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGG AGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACTGTCAGCTTTCTGATTG GTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCCTCACAGGAAAAATGATCCGAAGACGAACAGTG ACCCAGCCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCAGTCCAAACCCTGCCCAG TGAAGCCTTGTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGG AGAAGGGACCAGAACAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAA GTGGTGGATGAGGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGG AGGAATCCTGCAGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGAGCCTGTG TCAGCTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCAGATCCAGACCGGTG ATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAATCATGTTATG ATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTTCCCGAACAGTGTGGTG TCAAAGGTCAGATGGTATAAATGTAACAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCT GGGGCATTTGTGTTACTCATCTTTATTGTCTCCATGATTTATCTAGCTTGCAAAAAGCCAAAGAAAC CCCAAAGAAGGCAAAACAACCGACTGAAACCTTTAACCTTAGCCTATGATGGAGATGCCGACATGTA ACATATAACTTTTCCTGGCAACAACCAGTTTCGGCTTTCTGACTTCATAGATGTCCAGAGGCCACAA CAAATGTATCCAAACTGTGTGGATTAAAATATATTTTAATTTTAAAAATGGCATCATAAAGACAAG AGTGAAAATCATACTGCCACTGGAGATATTTAAGACAGTACCACTTATATACAGAT		
	ORF Start: ATG at 18		ORF Stop: TAA at 4488

	SEQ ID NO: 184	1490 aa	MW at 167403.2kD
NOV47e, CG51595-07 Protein Sequence	MGDECGPGGIQTRAVWCAHVEGWTTLHTNCKQAERPNNQQNCFKVCDWHKELYDWRLGPWNQCQPVI SKSLEKPLECIKGEEGIQVREIACIQKDKDIPAEDIICEYFEPKPLLEQACLIPCQQDCIVSEFSAW SECSKTCGSGLQHRTRHVVAPPQFGGSGCPNLTEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHSRQ VRQARRRGKNKEREKDRSKGVKDPEARELIKKKRNRNRQNRQENKYWDIQIGYQTREVMCINKTGKA ADLSFCQQEKLPMTFQSCVITKECQVSEWSEWSPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKECP EFEEKEPCLSQGDGVVPCATYGWRTTEWTECRVDPLLSQQDKRRGNQTALCGGIQTREVYCVQANE NLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQLCHIPCPTECEVSPWSAWGPCTYENCNDQQGKKGF KLRKRRITNEPTGGSGVTGNCPHLLEAIPCEEPACYDWKAVRLGDCEPDNGKECGPGTQVQEVVCIN SDGEEVDRQLCRDAIFPIPVACDAPCPKDCVLSTWSTWSSCSHTCSGKTTEGKQIRARSILAYAGEE GGIRCPNSSALQEVRSCNEHPCTVYHWQTGPWGQCIEDTSVSSFNTTTTWNGEASCSVGMQTRKVIC		

	VRVNVGQVGPKKCPESLRPETVRPCLLPCKKDCIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVIIQL PANGGRDCTDPLYEEKACEAPQACQSYRWKTHKWRRCQLVPWSVQQDSPGAQEGCGPGRQARAITCR KQDGGQAGIHECLQYAGPVPALTQACQIPCQDDCQLTSWSKFSSCNGDCGAVRTRKRTLVGKSKKKE KCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLLGMKVQGDIKECGQGYRYQAMACYD QNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKSCGSGVKVRSKWLREKPYNGGRPCP KLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMRENCGEGVQTRKVRCMQNTADGPSEHV EDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQSSFRQRSADPIRQPADEGRSCPNAV EKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDCVRSDGKSVDLKYCEALGLEKNWQM NTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPFQGDGRPCPSLMDQSKPCPVKPCYR WQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDEEFCADIELIIDGNKNMVLEESCSQ PCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQELENQHLCPEQMLETKSCYDGQCYE YKWMASAWKGSSRTVWCQRSDGINVTDGRLKTWVYGVAAGAFVLLIFIVSMIYLACKKPKKPQRRQN NRLKPLTLAYDGDADM

	SEQ ID NO: 185	4647 bp
NOV47f, 306395637 DNA Sequence	GGTACCATGGGAGATGAATGTGGTCCCGGAGGCATCCAAACGAGGGCTGTGTGGTGTGCTCATGTGG AGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGAGAGACCCAATAACCAGCAGAATTGTTT CAAAGTTTGCGATTGGCACAAAGAGTTGTACGACTGGAGACTGGGACCTTGGAATCAGTGTCAGCCC GTGATTTCAAAAGCCTAGAGAAACCTCTTGAGTGCATTAAGGGGGAAGAAGGTATTCAGGTGAGGG AGATAGCGTGCATCCAGAAAGACAAAGACATTCCTGCGGAGGATATCATCTGTGAGTACTTTGAGCC CAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTTGCCAGCAAGATTGCATCGTGTCTGAATTTTCT GCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGGGCTCCAGCACCGGACGCGTCATGTGGTGGCGC CCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAGGTGTGCCAATCCAGTCCATG CGAGGCCGAGGAGCTCAGGTACAGCCTCGCATGTGGGGCCCTGGAGCACCTGCTCAATGCCCCACTCC CGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAGAACGGGAAAAGGACCGCAGCAAAGGAG TAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGAGAAACAGAAACAGACAGAACAGACAAGA GAACAAATATTGGGACATCCAGATTGGATATCAGACCAGAGAGGTTATGTGCATTAACAAGACGGGG AAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAAGCTTCCAATGACCTTCCAGTCCTGTGTGATCA CCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGCGGAGCCCCTGCTCAAAAACATGCCATGACATGGT GTCCCCTGCAGGCACTCGTGTAAGGACACGAACCATCAGGCAGTTTCCCATTGGCAGTGAAAAGGAG TGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAGGAGATGGAGTTGTCCCCTGTGCCACGT ATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGGACCCTTTGCTCAGTCAGCAGGACAAGAG GCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAGACCCGAGAGGTGTACTGCGTGCAGGCC AACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACAAAGAAGCCTCAAAGCCAATGGACTTAA AATTATGCACTGGACCTATCCCTAATACTACACAGCTGTGCCACATTCCTTGTCCAACTGAATGTGA AGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTTATGAAAACTGTAATGATCAGCAAGGGAAAAAA GGCTTCAAACTGAGGAAGCGGCGCATTACCAATGAGCCCACTGGAGGCTCTGGGGTAACCGGAAACT GCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAGCCTGCCTGTTATGACTGGAAAGCAGTGAGACT GGGAAACTGCGAGCCAGATAACGGAAAGGAGTGTGGTCCAGGCACGCAAGTTCAAGAGGTTGTGTGC ATCAACAGTGATGGAGAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCCCCATCCCTGTGG CCTGTGATGCCCCGTGCCCGAAAGACTGTGTGCTCAGCACATGGTCTACGTGGTCCTCCTGCTCACA CACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGAGCACGATCCATTCTGGCCTATGCGGGT GAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGCAAGAAGTACGAAGCTGTAATGAGCATC CTTGCACAGTGTACCACTGGCAAACTGGTCCCTGGGGCCAGTGCATTGAGGACACCTCAGTATCGTC CTTCAACACAACTACGACTTGGAATGGGGAGGCCTCCTGCTCTGTCGGCATGCAGACAAGAAAAGTC ATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAAAATGTCCTGAAAGCCTTCGACCTGAAA CTGTAAGGCCTTGTCTGCTTCCTTGTAAGAAGGAGTGTATTGTGACCCCATATAGTGACTGGACATC ATGCCCCTCTTCGTGTAAAGAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCATCGGGTCATCATT CAGCTGCCAGCCAACGGGGGCCGAGACTGCACAGATCCCCTCTATGAAGAGAAGGCCTGTGAGGCAC CTCAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATGGCGCAGATGCCAATTAGTCCCTTGGAG CGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGGCCTGGGCGACAGGCAAGAGCCATTACT TGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGCAGGCCCTGTGCCAG CCCTTACCCAGGCCTGCCAGATCCCCTGCCAGGATGACTGTCAATTGACCAGCTGGTCCAAGTTTTC TTCATGCAATGGAGACTGTGGTGCAGTTAGGACCAGAAAGCGCACTCTTGTTGGAAAAAGTAAAAAG AAGGAAAAATGTAAAAATTCCCATTTGTATCCCCTGATTGAGACTCAGTATTGTCCTTGTGACAAAT ATAATGCACAACCTGTGGGGAACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTGGAAGTGTTGCT GGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAGCAATGGCATGC TACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACATTGAGGAGGCCT GCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACTGGTCGCGCTGCAGCAAGTC CTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTGAAAAACCATATAATGGAGGAAGGCCT TGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATGAGGTTGTCCCATGCCACAGTGACTGCA ACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAGAA CTGTGGAGAGGGCGTGCAAACCCGAAAAGTGAGATGCATGCAGAATACAGCAGATGGCCCTTCTGAA CATGTAGAGGATTACCTCTGTGACCCAGAAGAGATGCCCCTGGGCTCTAGAGTGTGCAAATTACCAT GCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCTTGCAATCAAAG CAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGATCTTGCCCTAAT GCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATTATAATGTAACAGACT GGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAATAAAAACAAGGATGTTGGATTG TGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAGAACTGG CAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACTGTCAGCTTTCTGATTGGTCTCCTTGGT CAGAATGTTCTCAAACATGTGGCCTCACAGGAAAAATGATCCGAAGACGAACAGTGACCCAGCCCTT TCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCAGTCCAAACCCTGCCCAGTGAAGCCTTGT TATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGGAGAAGGGACCA	

	GAACAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAAGTGGTGGATGA GGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGGAGGAATCCTGC AGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGAGCCTGTGTCAGCTGACCT GTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCAGATCCAGACCGGTGATTATACAAGA ACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAATCATGTTATGATGGACAGTGC TATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTTCCCGAACAGTGTGGTGTCAAAGGTCAG ATGGTATAAATGTAACAGGGGGCTGCTTGGTGATGAGCCAGCCTGATGCCGACAGGTCTTGTAACCC ACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAAACATGCCATTGTGAAGAAGGGTACACT GAAGTCATGTCTTCTAACAGCACCCTTGAGCAATGCACACTTATCCCCGTGGTGGTATTACCCACCA TGGAGGACAAAAGAGGAGATGTGAAAACCAGTCGGGCTGTACATCCAACCCAACCCTCCAGTAACCC AGCAGGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGGCCAGCAAAAAGCCAAAGAAACCCCAAA GAAGGCAAAACAACCGACGTCGAC	
	ORF Start: at 1	ORF Stop: end of sequence

	SEQ ID NO: 186	1549 aa	MW at 173501.6kD
NOV47f, 306395637 Protein Sequence	GTMGDECGPGGIQTRAVWCAHVEGWTTLHTNCKQAERPNNQQNCFKVCDWHKELYDWRLGPWNQCQP VISKSLEKPLECIKGEEGIQVREIACIQKDKDIPAEDIICEYFEPKPLLEQACLIPCQQDCIVSEFS AWSECSKTCGSGLQHRTRHVVAPPQFGGSGCPNLTEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHS RQVRQARRRGKNKEREKDRSKGVKDPEARELIKKKRNRNRQNRQENKYWDIQIGYQTREVMCINKTG KAADLSFCQQEKLPMTFQSCVITKECQVSEWSERSPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKE CPEFEEKEPCLSQGDGVVPCATYGWRTTEWTECRVDPLLSQQDKRRGNQTALCGGGIQTREVYCVQA NENLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQLCHIPCPTECEVSPWSAWGPCTYENCNDQQGKK GFKLRKRRITNEPTGGSGVTGNCPHLLEAIPCEEPACYDWKAVRLGNCEPDNGKECGPGTQVQEVVC INSDGEEVDRQLCRDAIFPIPVACDAPCPKDCVLSTWSTWSSCSHTCSGKTTEGKQIRARSILAYAG EEGGIRCPNSSALQEVRSCNEHPCTVYHWQTGPWGQCIEDTSVSSFNTTTTWNGEASCSVGMQTRKV ICVRVNVGQVGPKKCPESLRPETVRPCLLPCKKECIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVII QLPANGGRDCTDPLYEEKACEAPQACQSYRWKTHKWRRCQLVPWSVQQDSPGAQEGCGPGRQARAIT CRKQDGGQAGIHECLQYAGPVPALTQACQIPCQDDCQLTSWSKFSSCNGDCGAVRTRKRTLVGKSKK KEKCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLLGMKVQGDIKECGQGYRYQAMAC YDQNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKSCGSGVKVRSKWLREKPYNGGRP CPKLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMRENCGEGVQTRKVRCMQNTADGPSE HVEDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQSSFRQRSADPIRQPADEGRSCPN AVEKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDCVRSDGKSVDLKYCEALGLEKNW QMNTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPFQGDGRPCPSLMDQSKPCPVKPC YRWQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDEEFCADIELIIDGNKNMVLEESC SQPCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQELENQHLCPEQMLETKSCYDGQC YEYKWMASAWKGSSRTVWCQRSDGINVTGGCLVMSQPDADRSCNPPCSQPHSYCSETKTCHCEEGYT EVMSSNSTLEQCTLIPVVVLPTMEDKRGDVKTSRAVHPTQPSSNPAGRGRTWFLQPFGPAKSQRNPK EGKTTDVD		

	SEQ ID NO: 187	6373 bp	
NOV47g, CG51595-01 DNA Sequence	GACAGAGTGCAGCCTTTTCAGACTCTGTGACACAGTTCCCCTTTTGCAAAAATACTTAGCGAGGATC ATTACTTTCCAACAGTCGTGTCCAGAGACCTACTTTGTAACACCGCAGGGAAGTTAATGTACTAGGT CTTGAAAGGTCTTTCTGGAATGTGCAGTAACTTGTAGTTTTCTTCTAGTAGCACTGCTAATTTTTGT GTTATAATTTTTGTAGGTCCATGGGGCCGATGTATGGAGATGAATGTGGTCCCGGAGGCATCCAAA CGAGGGCTGTGTGGTGTGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGA GAGACCCAATAACCAGCAGAATTGTTTCAAAGTTTGCGATTGGCACAAAGAGTTGTACGACTGGAGA CTGGGACCTTGGAATCAGTGTCAGCCCGTGATTTCAAAAAGCCTAGAGAAACCTCTTGAGTGCATTA AGGGGGAAGAAGGTATTCAGGTGAGGGAGATAGCGTGCATCCAGAAAGACAAAGACATTCCTGCGGA GGATATCATCTGTGAGTACTTTGAGCCCAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTTGCCAG CAAGATTGCATCGTGTCTGAATTTTCTGCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGGGCTCC AGCACCGGACGCGTCATGTGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGA GTTCCAGGTGTGCCAATCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCC TGGAGCACCTGCTCAATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAG AACGGGAAAAGGACCGCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGAG AAACAGAAACAGGCAGAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATCAGACCAGA GAGGTTATGTGCATTAACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAAGCTTC CAATGACCTTCCAGTCCTGTGTGATCACCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGTGGAGCCC CTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACACGAACCATCAGG CAGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAG GAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGGA CCCTTTGCTCAGTCAGCAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAG ACCCGAGAGGTGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACA AAGAAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGCTGTG CCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTTATGAA AACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAATGAGCCCA CTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAGCCTGC CTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGGAGTGTGGTCCA GGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAGACAGCTGTGCA		

	GAGATGCCATCTTCCCCATCCCTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTCAGCAC ATGGTCTACGTGGTCCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGA GCACGATCCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGC AAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCTGGGGCCA GTGCATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATGGGGAGGCCTCCTGC TCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGACCCAAAA AATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCTTGTAAGAAGGACTGTAT TGTGACCCCATATAGTGACTGGACATCATGCCCCTCTTCGTGTAAAGAAGGGGACTCCAGTATCAGG AAGCAGTCTAGGCATCGGGTCATCATTCAGCTGCCAGCCAACGGGGGCCGAGACTGCACAGATCCCC TCTATGAAGAGAAGGCCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATG GCGCAGATGCCAATTAGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGG CCTGGGCGACAGGCAAGAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATGAGT GCCTACAGTATGCAGGCCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCCAGGATGACTG TCAATTGACCAGCTGGTCCAAGTTTTCTTCATGCAATGGAGACTGTGGTGCAGTTAGGACCAGAAAG CGCACTCTTGTTGGAAAAAGTAAAAAGAAGGAAAAATGTAAAAATTCCCATTTGTATCCCCTGATTG AGACTCAGTATTGTCCTTGTGACAAATATAATGCACAACCTGTGGGGAACTGGTCAGACTGTATTTT ACCAGAGGGAAAAGTGGAAGTGTTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAA GGATATCGTTACCAAGCAATGGCATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTA ACAGCCATGGTTACATTGAGGAGGCCTGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTG GTCCAACTGGTCGCGCTGCAGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGT GAAAAACCATATAATGGAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATG AGGTTGTCCCATGCCACAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAA GGTGACCTTTGTGAATATGCGGGAGAACTGTGGAGAGGGCGTGCAAACCCGAAAAGTGAGATGCATG CAGAATACAGCAGATGGCCCTTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAAGAGATGCCCC TGGGCTCTAGAGTGTGCAAATTACCATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGAC CCAATGTGTTTTGCCTTGCAATCAAAGCAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCA GCTGATGAAGGAAGATCTTGCCCTAATGCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCT ACCACTATGATTATAATGTAACAGACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAA TGGAATAAAAACAAGGATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGT GAAGCGCTTGGCTTGGAGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACT GTCAGCTTTCTGATTGGTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCCTCACAGGAAAAATGAT CCGAAGACGAACAGTGACCCAGCCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCAG TCCAAACCCTGCCCAGTGAAGCCTTGTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGC AGGAGGCCCAGTGTGGAGAAGGGACCAGAACAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGC TGATGATTTCAGCAAAGTGGTGGATGAGGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAAT AAAAATATGGTTCTGGAGGAATCCTGCAGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGT CTTCCTGGAGCCTGTGTCAGCTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGT CAGATCCAGACCGGTGATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAA ACAAAATCATGTTATGATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTT CCCGAACAGTGTGGTGTCAAAGGTCAGATGGTATAAATGTAACAGGGGGCTGCTTGGTGATGAGCCA GCCTGATGCCGACAGGTCTTGTAACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAA ACATGCCATTGTGAAGAAGGGTACACTGAAGTCATGTCTTCTAACAGCACCCTTGAGCAATGCACAC TTATCCCCGTGGTGGTATTACCCACCATGGAGGACAAAAGAGGAGATGTGAAAACCAGTCGGGCTGT ACATCCAACCCAACCCTCCAGTAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGG CCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTTGTGTTACTCATCTTTA TTGTCTCCATGATTTATCTAGCTTGCAAAAAGCCAAAGAAACCCCAAAGAAGGCAAAACAACCGACT GAAACCTTTAACCTTAGCCTATGATGGAGATGCCGACATG**TAA**CATATAACTTTTCCTGGCAACAAC CAGTTTCGGCTTTCTGACTTCATAGATGTCCAGAGGCCACAACAAATGTATCCAAACTGTGTGGATT AAAATATATTTTAATTTTTAAAAATGGCATCATAAAGACAAGAGTGAAAATCATACTGCCACTGGAG ATATTTAAGACAGTACCACTTATATACAGACCATCAACCGTGAGAATTATAGGAGATTTAGCTGAAT ACATGCTGCATTCTGAAAGTTTTATGTCATCTTTTCTGAAATCTACCGACTGAAAAACCACTTTCAT CTCTAAAAAATAATGGTGGAATTGGCCAGTTAGGATGCCTGATACAAGACCGTCTGCAGTGTTAATC CATAAAACTTCCTAGCATGAAGAGTTTCTACCAAGATCTCCACAATACTATGGTCAAATTAACATGT GTACTCAGTTGAATGACACACATTATGTCAGATTATGTACTTGCTAATAAGCAATTTTAACAATGCA TAACAAATAAACTCTAAGCTAAGCAGAAAATCCACTGAATAAATTCAGCATCTTGGTGGTCGATGGT AGATTTTATTGACCTGCATTTCAGAGACAAAGCCTCTTTTTTAAGACTTCTTGTCTCTCTCCAAAGT AAGAATGCTGGACAAGTACTAGTGTCTTAGAAGAACGAGTCCTCAAGTTCAGTATTTTATAGTGGTA ATTGTCTGGAAAACTAATTTACTTGTGTTAATACAATACGTTTCTACTTTCCCTGATTTTCAAACTG GTTGCCTGCATCTTTTTTGCTATATGGAAGGCACATTTTTGCACTATATTAGTGCAGCACGATAGGC GCTTAACCAGTATTGCCATAGAAACTGCCTCTTTTCATGTGGATGAAGACATCTGTGCCAAGAGTG GCATGAAGACATTTGCAAGTTCTTGTATCCTGAAGAGAGTAAAGTTCAGTTTGGATGGCAGCAAGAT GAAATCAGCTATTACACCTGCTGTACACACACTTCCTCATCACTGCAGCCATTGTGAAATTGACAAC ATGGCGGTAATTTAAGTGTTGAAGTCCCTAACCCCTTAACCCTCTAAAAGGTGGATTCCTCTAGTTG GTTTGTAATTGTTCTTTGAAGGCTGTTTATGACTAGATTTTTATATTTGTTATCTTTGTTAAGAAAA AAAAAAGAAAAAGGAACTGGATGTCTTTTTAATTTTGAGCAGATGGAGAAAATAAATAATGTATCAA TGACCTTTGTAACTAAAGGAAAAAAAAAAAAAATGTGGATTTTCCTTTCTCTCTGATTTCCCAGTTT CAGATTGAATGTCTGTCTTGCAGGCAGTTATTTCAAAATCCATAGTCTTTNGCCTTTCTCACTGGCA AAATTTGA

	ORF Start: ATG at 235		ORF Stop: TAA at 4999

	SEQ ID NO: 188	1588 aa	MW at 178042.1kD

NOV47g, CG51595-01 Protein Sequence	MGDECGPGGIQTRAVWCAHVEGWTTLHTNCKQAERPNNQQNCFKVCDWHKELYDWRLGPWNQCQPVI SKSLEKPLECIKGEEGIQVREIACIQKDKDIPAEDIICEYFEPKPLLEQACLIPCQQDCIVSEFSAW SECSKTCGSGLQHRTRHVVAPPQFGGSGCPNLTEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHSRQ VRQARRRGKNKEREKDRSKGVKDPEARELIKKKRNRNRQNRQENKYWDIQIGYQTREVMCINKTGKA ADLSFCQQEKLPMTFQSCVITKECQVSEWSEWSPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKECP EFEEKEPCLSQGDGVVPCATYGWRTTEWTECRVDPLLSQQDKRRGNQTALCGGGIQTREVYCVQANE NLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQLCHIPCPTECEVSPWSAWGPCTYENCNDQQGKKGF KLRKRRITNEPTGGSGVTGNCPHLLEAIPCEEPACYDWKAVRLGDCEPDNGKECGPGTQVQEVVCIN SDGEEVDRQLCRDAIFPIPVACDAPCPKDCVLSTWSTWSSCSHTCSGKTTEGKQIRARSILAYAGEE GGIRCPNSSALQEVRSCNEHPCTVYHWQTGPWGQCIEDTSVSSFNTTTTWNGEASCSVGMQTRKVIC VRVNVGQVGPKKCPESLRPETVRPCLLPCKKDCIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVIIQL PANGGRDCTDPLYEEKACEAPQACQSYRWKTHKWRRCQLVPWSVQQDSPGAQEGCGPGRQARAITCR KQDGGQAGIHECLQYAGPVPALTQACQIPCQDDCQLTSWSKFSSCNGDCGAVRTRKRTLVGKSKKKE KCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLLGMKVQGDIKECGQGYRYQAMACYD QNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKSCGSGVKVRSKWLREKPYNGGRPCP KLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMRENCGEGVQTRKVRCMQNTADGPSEHV EDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQSSFRQRSADPIRQPADEGRSCPNAV EKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDCVRSDGKSVDLKYCEALGLEKNWQM NTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPFQGDGRPCPSLMDQSKPCPVKPCYR WQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDEEFCADIELIIDGNKNMVLEESCSQ PCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQELENQHLCPEQMLETKSCYDGQCYE YKWMASAWKGSSRTVWCQRSDGINVTGGCLVMSQPDADRSCNPPCSQPHSYCSETKTCHCEEGYTEV MSSNSTLEQCTLIPVVVLPTMEDKRGDVKTSRAVHPTQPSSNPAGRGRTWFLQPFGPDGRLKTWVYG VAAGAFVLLIFIVSMIYLACKKPKKPQRRQNNRLKPLTLAYDGDADM

	SEQ ID NO: 189	1605 bp
NOV47h, 283842727 DNA Sequence	GGTACCGATATCATCTGTGAGTACTTTGAGCCCAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTT GCCAGCAAGATTGCATCGTGTCTGAATTTTCTGCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGG GCTCCAGCACCGGACGCGTCATGTGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTG ACGGAGTTCCAGGTGTGCCAATCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGG GGCCCTGGAGCACCTGCTCAATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAA TAAAGAACGGGAAAAGGACCGCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAA AAGAGAAACAGAAACAGGCAGAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATCAGA CCAGAGAGGTTATGTGCATTAACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAA GCTTCCAATGACCTTCCAGTCCTGTGTGATCACCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGTGG AGCCCCTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACACGAACCA TCAGGCAGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTC TCAAGGAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGT GTGGACCCTTTGCTCAGTCAGCAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCA TCCAGACCCGAGAGGTGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAA GAACAAAGAAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAG CTGTGCCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTT ATGAAAACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAATGA GCCCACTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAG CCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGGAGTGTG GTCCAGGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAGACAGCT GTGCAGAGATGCCATCTTCCCCATCCCTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTC AGCACATGGTCTACGTGGTCCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGA TACGAGCACGATCCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGC TTTGCAAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGTCGAC	
	ORF Start: at 1	ORF Stop: end of sequence

	SEQ ID NO: 190	535 aa	MW at 59956.1kD
NOV47h, 283842727 Protein Sequence	GTDIICEYFEPKPLLEQACLIPCQQDCIVSEFSAWSECSKTCGSGLQHRTRHVVAPPQFGGSGCPNL TEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHSRQVRQARRRGKNKEREKDRSKGVKDPEARELIKK KRNRNRQNRQENKYWDIQIGYQTREVMCINKTGKAADLSFCQQEKLPMTFQSCVITKECQVSEWSEW SPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKECPEFEEKEPCLSQGDGVVPCATYGWRTTEWTECR VDPLLSQQDKRRGNQTALCGGGIQTREVYCVQANENLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQ LCHIPCPTECEVSPWSAWGPCTYENCNDQQGKKGFKLRKRRITNEPTGGSGVTGNCPHLLEAIPCEE PACYDWKAVRLGDCEPDNGKECGPGTQVQEVVCINSDGEEVDRQLCRDAIFPIPVACDAPCPKDCVL STWSTWSSCSHTCSGKTTEGKQIRARSILAYAGEEGGIRCPNSSALQEVRSCNEHPCTVYHWQTVD		

	SEQ ID NO: 191	1605 bp
NOV47i,	GGTACCGATATCATCTGTGAGTACTTTGAGCCCAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTT GCCAGCAAGATTGCATCGTGTCTGAATTTTCTGCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGG	

283842704 DNA Sequence	GCTCCAGCACCGGACGCGTCATGTGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTG ACGGAGTTCCAGGTGTGCCAATCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGG GGCCCTGGAGCACCTGCTCAATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAA TAAAGAACGGGAAAAGGACCGCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAA AAGAGAAACAGAAACAGGCAGAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATCAGA CCAGAGAGGTTATGTGCATTAACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAA GCTTCCAATGACCTTCCAGTCCTGTGTGATCACCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGTGG AGCCCCTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACACGAACCA TCAGGCAGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTC TCAAGGAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGT GTGGACCCTTTGCTCAGTCAGCAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCA TCCAGACCCGAGAGGTGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAA GAACAAAGAAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAG CTGTGCCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTT ATGAAAACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAATGA GCCCACTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAG CCTGCCTGTTATGACTGGAAAGCAGTGAGACTGGGAAACTGCGAGCCAGATAACGGAAAGGAGTGTG GTCCAGGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAGACAGCT GTGCAGAGATGCCATCTTCCCCATCCCTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTC AGCACATGGTCTACGTGGTCCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGA TACGAGCACGATCCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGC TTTGCAAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGTCGAC	
	ORF Start: at 1	ORF Stop: end of sequence

	SEQ ID NO: 192	535 aa	MW at 59955.1kD
NOV47i, 283842704 Protein Sequence	GTDIICEYFEPKPLLEQACLIPCQQDCIVSEFSAWSECSKTCGSGLQHRTRHVVAPPQFGGSGCPNL TEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHSRQVRQARRRGKNKEREKDRSKGVKDPEARELIKK KRNRNRQNRQENKYWDIQIGYQTREVMCINKTGKAADLSFCQQEKLPMTFQSCVITKECQVSEWSEW SPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKECPEFEEKEPCLSQGDGVVPCATYGWRTTEWTECR VDPLLSQQDKRRGNQTALCGGGIQTREVYCVQANENLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQ LCHIPCPTECEVSPWSAWGPCTYENCNDQQGKKGFKLRKRRITNEPTGGSGVTGNCPHLLEAIPCEE PACYDWKAVRLGNCEPDNGKECGPGTQVQEVVCINSDGEEVDRQLCRDAIFPIPVACDAPCPKDCVL STWSTWSSCSHTCSGKTTEGKQIRARSILAYAGEEGGIRCPNSSALQEVRSCNEHPCTVYHWQTVD		

	SEQ ID NO: 193	6373 bp	
NOV47j, CG51595-01 DNA Sequence	GACAGAGTGCAGCCTTTTCAGACTCTGTGACACAGTTCCCCTTTTGCAAAAATACTTAGCGAGGATC ATTACTTTCCAACAGTCGTGTCCAGAGACCTACTTTGTAACACCGCAGGGAAGTTAATGTACTAGGT CTTGAAAGGTCTTTCTGGAATGTGCAGTAACTTGTAGTTTTCTTCTAGTAGCACTGCTAATTTTTGT GTTATAATTTTTGTAGGTCCATGGGGCCGATGT**ATG**GGAGATGAATGTGGTCCCGGAGGCATCCAAA CGAGGGCTGTGTGGTGTGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGA GAGACCCAATAACCAGCAGAATTGTTTCAAAGTTTGCGATTGGCACAAAGAGTTGTACGACTGGAGA CTGGGACCTTGGAATCAGTGTCAGCCCGTGATTTCAAAAAGCCTAGAGAAACCTCTTGAGTGCATTA AGGGGGAAGAAGGTATTCAGGTGAGGGAGATAGCGTGCATCCAGAAAGACAAAGACATTCCTGCGGA GGATATCATCTGTGAGTACTTTGAGCCCAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTTGCCAG CAAGATTGCATCGTGTCTGAATTTTCTGCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGGGCTCC AGCACCGGACGCGTCATGTGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGA GTTCCAGGTGTGCCAATCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCC TGGAGCACCTGCTCAATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAG AACGGGAAAAGGACCGCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGAG AAACAGAAACAGGCAGAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATCAGACCAGA GAGGTTATGTGCATTAACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAAGCTTC CAATGACCTTCCAGTCCTGTGTGATCACCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGTGGAGCCC CTGCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACACGAACCATCAGG CAGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAG GAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGGA CCCTTTGCTCAGTCAGCAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCCAG ACCCGAGAGGTGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGAACA AAGAAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGCTGTG CCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTTATGAA AACTGTAATGATCAGCAAGGGAAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAATGAGCCCA CTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAGCCTGC CTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGGAGTGTGGTCCA GGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAGACAGCTGTGCA GAGATGCCATCTTCCCCATCCCTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTGTGCTCAGCAC ATGGTCTACGTGGTCCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGAAACAGATACGA GCACGATCCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATAGCAGTGCTTTGC AAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTGGTCCCTGGGGCCA GTGCATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATGGGGAGGCCTCCTGC TCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGACCCAAAA		

	AATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCTTGTAAGAAGGACTGTAT TGTGACCCCATATAGTGACTGGACATCATGCCCCTCTTCGTGTAAAGAAGGGGACTCCAGTATCAGG AAGCAGTCTAGGCATCGGGTCATCATTCAGCTGCCAGCCAACGGGGGCCGAGACTGCACAGATCCCC TCTATGAAGAGAAGGCCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAGACTCACAAATG GCGCAGATGCCAATTAGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGAGCACAGGAAGGCTGTGGG CCTGGGCGACAGGCAAGAGCCATTACTTGTCGCAAGCAAGATGGAGGACAGGCTGGAATCCATGAGT GCCTACAGTATGCAGGCCCTGTGCCAGCCCTTACCCAGGCCTGCCAGATCCCCTGCCAGGATGACTG TCAATTGACCAGCTGGTCCAAGTTTTCTTCATGCAATGGAGACTGTGGTGCAGTTAGGACCAGAAAG CGCACTCTTGTTGGAAAAAGTAAAAAGAAGGAAAAATGTAAAAATTCCCATTTGTATCCCCTGATTG AGACTCAGTATTGTCCTTGTGACAAATATAATGCACAACCTGTGGGGAACTGGTCAGACTGTATTTT ACCAGAGGGAAAAGTGGAAGTGTTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAA GGATATCGTTACCAAGCAATGGCATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTA ACAGCCATGGTTACATTGAGGAGGCCTGCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTG GTCCAACTGGTCGCGCTGCAGCAAGTCCTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGT GAAAAACCATATAATGGAGGAAGGCCTTGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATG AGGTTGTCCCATGCCACAGTGACTGCAACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAA GGTGACCTTTGTGAATATGCGGGAGAACTGTGGAGAGGGCGTGCAAACCCGAAAAGTGAGATGCATG CAGAATACAGCAGATGGCCCTTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAAGAGATGCCCC TGGGCTCTAGAGTGTGCAAATTACCATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGAC CCAATGTGTTTTGCCTTGCAATCAAAGCAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCA GCTGATGAAGGAAGATCTTGCCCTAATGCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCT ACCACTATGATTATAATGTAACAGACTGGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAA TGGAATAAAAACAAGGATGTTGGATTGTGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGT GAAGCGCTTGGCTTGGAGAAGAACTGGCAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACT GTCAGCTTTCTGATTGGTCTCCTTGGTCAGAATGTTCTCAAACATGTGGCCTCACAGGAAAAATGAT CCGAAGACGAACAGTGACCCAGCCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCAG TCCAAACCCTGCCCAGTGAAGCCTTGTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGC AGGAGGCCCAGTGTGGAGAAGGGACCAGAACAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGC TGATGATTTCAGCAAAGTGGTGGATGAGGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAAT AAAAATATGGTTCTGGAGGAATCCTGCAGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGT CTTCCTGGAGCCTGTGTCAGCTGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGT CAGATCCAGACCGGTGATTATACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAA ACAAAATCATGTTATGATGGACAGTGCTATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTT CCCGAACAGTGTGGTGTCAAAGGTCAGATGGTATAAATGTAACAGGGGGCTGCTTGGTGATGAGCCA GCCTGATGCCGACAGGTCTTGTAACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAA ACATGCCATTGTGAAGAAGGGTACACTGAAGTCATGTCTTCTAACAGCACCCCTTGAGCAATGCACAC TTATCCCCGTGGTGGTATTACCCACCATGGAGGACAAAAGAGGAGATGTGAAAACCAGTCGGGCTGT ACATCCAACCCAACCCTCCAGTAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGG CCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTTGTGTTACTCATCTTTA TTGTCTCCATGATTTATCTAGCTTGCAAAAAGCCAAAGAAACCCCAAAGAAGGCAAAACAACCGACT GAAACCTTTAACCTTAGCCTATGATGGAGATGCCGACATGTAACATATAACTTTTCCTGGCAACAAC CAGTTTCGGCTTTCTGACTTCATAGATGTCCAGAGGCCACAACAAATGTATCCAAACTGTGTGGATT AAAATATATTTTAATTTTTAAAAATGGCATCATAAAGACAAGAGTGAAAATCATACTGCCACTGGAG ATATTTAAGACAGTACCACTTATATACAGACCATCAACCGTGAGAATTATAGGAGATTTAGCTGAAT ACATGCTGCATTCTGAAAGTTTTATGTCATCTTTTCTGAAATCTACCGACTGAAAAACCACTTTCAT CTCTAAAAAATAATGGTGGAATTGGCCAGTTAGGATGCCTGATACAAGACCGTCTGCAGTGTTAATC CATAAAACTTCCTAGCATGAAGAGTTTCTACCAAGATCTCCACAATACTATGGTCAAATTAACATGT GTACTCAGTTGAATGACACACATTATGTCAGATTATGTACTTGCTAATAAGCAATTTTAACAATGCA TAACAAATAAACTCTAAGCTAAGCAGAAAATCCACTGAATAAATTCAGCATCTTGGTGGTCGATGGT AGATTTTATTGACCTGCATTTCAGAGACAAAGCCTCTTTTTTAAGACTTCTTGTCTCTCTCCAAAGT AAGAATGCTGGACAAGTACTAGTGTCTTAGAAGAACGAGTCCTCAAGTTCAGTATTTTATAGTGGTA ATTGTCTGGAAAACTAATTTACTTGTGTTAATACAATACGTTTCTACTTTCCCTGATTTTCAAACTG GTTGCCTGCATCTTTTTTGCTATATGGAAGGCACATTTTTGCACTATATTAGTGCAGCACGATAGGC GCTTAACCAGTATTGCCATAGAAACTGCCTCTTTTCATGTGGGATGAAGACATCTGTGCCAAGAGTG GCATGAAGACATTTGCAAGTTCTTGTATCCTGAAGAGAGTAAAGTTCAGTTTGGATGGCAGCAAGAT GAAATCAGCTATTACACCTGCTGTACACACACTTCCTCATCACTGCAGCCATTGTGAAATTGACAAC ATGGCGGTAATTTAAGTGTTGAAGTCCCTAACCCCTTAACCCTCTAAAAGGTGGATTCCTCTAGTTG GTTTGTAATTGTTCTTTGAAGGCTGTTTATGACTAGATTTTTATATTTGTTATCTTTGTTAAGAAAA AAAAAAGAAAAAGGAACTGGATGTCTTTTTAATTTTGAGCAGATGGAGAAAATAAATAATGTATCAA TGACCTTTGTAACTAAAGGAAAAAAAAAAAAAATGTGGATTTTCCTTTCTCTCTGATTTCCCAGTTT CAGATTGAATGTCTGTCTTGCAGGCAGTTATTTCAAAATCCATAGTCTTTNGCCTTTCTCACTGGCA AAATTTGA		
	ORF Start: ATG at 235		ORF Stop: TAA at 4999

	SEQ ID NO: 194	1588 aa	MW at 178042.1kD
NOV47j, CG51595-01 Protein Sequence	MGDECGPGGIQTRAVWCAHVEGWTTLHTNCKQAERPNNQQNCFKVCDWHKELYDWRLGPWNQCQPVI SKSLEKPLECIKGEEGIQVREIACIQKDKDIPAEDIICEYFEPKPLLEQACLIPCQQDCIVSEFSAW SECSKTCGSGLQHRTRHVVAPPQFGGSGCPNLTEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHSRQ VRQARRRGKNKEREKDRSKGVKDPEARELIKKKRNRNRQNRQENKYWDIQIGYQTREVMCINKTGKA ADLSFCQQEKLPMTFQSCVITKECQVSEWSEWSPCSKTCHDMVSPAGTRVRTRTIRQFPIGSEKECP EFEEKEPCLSQGDGVVPCATYGWRTTEWTECRVDPLLSQQDKRRGNQTALCGGGIQTREVYCVQANE NLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQLCHIPCPTECEVSPWSAWGPCTYENCNDQQGKKGF		

	KLRKRRITNEPTGGSGVTGNCPHLLEAIPCEEPACYDWKAVRLGDCEPDNGKECGPGTQVQEVVCIN SDGEEVDRQLCRDAIFPIPVACDAPCPKDCVLSTWSTWSSCSHTCSGKTTEGKQIRARSILAYAGEE GGIRCPNSSALQEVRSCNEHPCTVYHWQTGPWGQCIEDTSVSSFNTTTTWNGEASCSVGMQTRKVIC VRVNVGQVGPKKCPESLRPETVRPCLLPCKKDCIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVIIQL PANGGRDCTDPLYEEKACEAPQACQSYRWKTHKWRRCQLVPWSVQQDSPGAQEGCGPGRQARAITCR KQDGGQAGIHECLQYAGPVPALTQACQIPCQDDCQLTSWSKFSSCNGDCGAVRTRKRTLVGKSKKKE KCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLLGMKVQGDIKECGQGYRYQAMACYD QNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKSCGSGVKVRSKWLREKPYNGGRPCP KLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMRENCGEGVQTRKVRCMQNTADGPSEHV EDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQSSFRQRSADPIRQPADEGRSCPNAV EKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDCVRSDGKSVDLKYCEALGLEKNWQM NTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPFQGDGRPCPSLMDQSKPCPVKPCYR WQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDEEFCADIELIIDGNKNMVLEESCSQ PCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQELENQHLCPEQMLETKSCYDGQCYE YKWMASAWKGSSRTVWCQRSDGINVTGGCLVMSQPDADRSCNPPCSQPHSYCSETKTCHCEEGYTEV MSSNSTLEQCTLIPVVVLPTMEDKRGDVKTSRAVHPTQPSSNPAGRGRTWFLQPFGPDGRLKTWVYG VAAGAFVLLIFIVSMIYLACKKPKKPQRRQNNRLKPLTLAYDGDADM

	SEQ ID NO: 195	1732 bp
NOV47k, 310658551 DNA Sequence	CACCTCGCGAGGAGACTGTGGTGCAGTTAGGACCAGAAAGCGCACTCTTGTTGGAAAAAGTAAAAAG AAGGAAAAATGTAAAAATTCCCATTTGTATCCCCTGATTGAGACTCAGTATTGTCCTTGTGACAAAT ATAATGCACAACCTGTGGGGAACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTGGAAGTGTTGCT GGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAGCAATGGCATGC TACGATCAAAATGGCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACATTGAGGAGGCCT GCATCATCCCCTGCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACTGGTCGCGCTGCAGCAAGTC CTGTGGGAGTGGTGTGAAGGTTCGTTCTAAATGGCTGCGTGAAAAACCATATAATGGAGGAAGGCCT TGCCCCAAACTGGACCATGTCAACCAGGCACAGGTGTATGAGGTTGTCCCATGCCACAGTGACTGCA ACCAGTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAGAA CTGTGGAGAGGGCGTGCAAACCCGAAAAGTGAGATGCATGCAGAATACAGCAGATGGCCCTTCTGAA CATGTAGAGGATTACCTCTGTGACCCAGAAGAGATGCCCCTGGGCTCTAGAGTGTGCAAATTACCAT GCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCTTGCAATCAAAG CAGTTTCCGGCAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGATCTTGCCCTAAT GCTGTTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATTATAATGTAACAGACT GGAGTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAATAAAAACAAGGATGTTGGATTG TGTTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAGAACTGG CAGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACTGTCAGCTTTCTGATTGGTCTCCTTGGT CAGAATGTTCTCAAACATGTGGCCTCACAGGAAAAATGATCCGAAGACGAACAGTGACCCAGCCCTT TCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCAGTCCAAACCCTGCCCAGTGAAGCCTTGT TATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGGAGAAGGGACCA GAACAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAAGTGGTGGATGA GGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGGAGGAATCCTGC AGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGAGCCTGTGTCAGCTGACCT GTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCAGATCCAGACCGGTGATTATACAAGA ACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAATCATGTTATGATGGACAGTGC TATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTTCCCGAACAGTCGACGGC	
	ORF Start: at 2	ORF Stop: end of sequence

	SEQ ID NO: 196	577 aa	MW at 65124.1kD
NOV47k, 310658551 Protein Sequence	TSRGDCGAVRTRKRTLVGKSKKKEKCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLL GMKVQGDIKECGQGYRYQAMACYDQNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKS CGSGVKVRSKWLREKPYNGGRPCPKLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMREN CGEGVQTRKVRCMQNTADGPSEHVEDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQS SFRQRSADPIRQPADEGRSCPNAVEKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDC VRSDGKSVDLKYCEALGLEKNWQMNTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPF QGDGRPCPSLMDQSKPCPVKPCYRWQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDE EFCADIELIIDGNKNMVLEESCSQPCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQE LENQHLCPEQMLETKSCYDGQCYEYKWMASAWKGSSRTVDG		

	SEQ ID NO: 197	921 bp
NOV47l, CG51595-02 DNA Sequence	ATGGGAGATGAATGTGGTCCCGGAGGCATCCAAACGAGGGCTGTGTGGTGTGCTCATGTGGAGGGAT GGACTACACTGCATACTAACTGTAAGCAGGCCGAGAGACCCAATAACCAGCAGAATTGTTTCAAAGT TTGCGATTGGCACAAAGAGTTGTACGACTGGAGACTGGGACCTTGGAATCAGTGTCAGCCCGTGATT TCAAAAAGCCTAGAGAAACCTCTTGAGTGCATTAAGGGGGAAGAAGGTATTCAGGTGAGGGAGATAG CGTGCATCCAGAAAGACAAAGACATTCCTGCGGAGGATATCATCTGTGAGTACTTTGAGCCCAAGCC TCTCCTGGAGCAGGCTTGCCTCATTCCTTGCCAGCAAGATTGCATCGTGTCTGAATTTTCTGCCTGG	

	TCCGAATGCTCCAAGACCTGCGGCAGCGGGCTCCAGCACCGGACGCGTCATGTGGTGGCGCCCCCGC AGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAGGTGTGCCAATCCAGTCCATGCGAGGC CGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCCTGGAGCACCTGCTCAATGCCCCACTCCCGACAA GTAAGACAAGCAAGGAGACGCGGGAAGAATAAAGAACGGGAAAAGGACCGCAGCAAAGGAGTAAAGG ATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGAGAAACAGAAACAGGCAGAACAGACAAGAGAACAA ATATTGGGACATCCAGATTGGATATCAGACCAGAGAGGTTATGTGCATTAACAAGACGGGGAAAGCT GCTGATTTAAGCTTTTGCCAGCAAGAGAAGCTTCCAATGACCTTCCAGTCCTGTGTGATCACCAAAG AGTGCCAGGTTTCCGAGTGGTCAGAGTGGAGCCCCTGCTCAAAAACATGC	
	ORF Start: ATG at 1	ORF Stop: end of sequence

	SEQ ID NO: 198	307 aa	MW at 35305.8kD
NOV47l, CG51595-02 Protein Sequence	MGDECGPGGIQTRAVWCAHVEGWTTLHTNCKQAERPNNQQNCFKVCDWHKELYDWRLGPWNQCQPVI SKSLEKPLECIKGEEGIQVREIACIQKDKDIPAEDIICEYFEPKPLLEQACLIPCQQDCIVSEFSAW SECSKTCGSGLQHRTRHVVAPPQFGGSGCPNLTEFQVCQSSPCEAEELRYSLHVGPWSTCSMPHSRQ VRQARRRGKNKEREKDRSKGVKDPEARELIKKKRNRNRQNRQENKYWDIQIGYQTREVMCINKTGKA ADLSFCQQEKLPMTFQSCVITKECQVSEWSEWSPCSKTC		

	SEQ ID NO: 199	4810 bp	
NOV47m, CG51595-05 DNA Sequence	GTCCATGGGCCGATGT**ATG**GGAGATGAATGTGGTCCCGGAGGCATCCAAACGAGGGCTGTGTGGT GTGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCCGAGAGACCCAATAACC AGCAGAATTGTTTCAAAGTTTGCGATTGGCACAAAGAGTTGTACGACTGGAGACTGGGACCTTGGA ATCAGTGTCAGCCCGTGATTTCAAAAAGCCTAGAGAAACCTCTTGAGTGCATTAAGGGGGAAGAAG GTATTCAGGTGAGGGAGATAGCGTGCATCCAGAAAGACAAAGACATTCCTGCGGAGGATATCATCT GTGAGTACTTTGAGCCCAAGCCTCTCCTGGAGCAGGCTTGCCTCATTCCTTGCCAGCAAGATTGCA TCGTGTCTGAATTTTCTGCCTGGTCCGAATGCTCCAAGACCTGCGGCAGCGGGCTCCAGCACCGGA CGCGTCATGTGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAGG TGTGCCAATCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCATGTGGGGCCCTGGAGCA CCTGCTCAATGCCCCACTCCCGACAAGTAAGACAAGCAAGGAGACGCGGGAAGAATAAAGAACGGG AAAAGGACCGCAGCAAAGGAGTAAAGGATCCAGAAGCCCGCGAGCTTATTAAGAAAAAGAGAAACA GAAACAGGCAGAACAGACAAGAGAACAAATATTGGGACATCCAGATTGGATATCAGACCAGAGAGG TTATGTGCATTAACAAGACGGGGAAAGCTGCTGATTTAAGCTTTTGCCAGCAAGAGAAGCTTCCAA TGACCTTCCAGTCCTGTGTGATCACCAAAGAGTGCCAGGTTTCCGAGTGGTCAGAGTGGAGCCCCT GCTCAAAAACATGCCATGACATGGTGTCCCCTGCAGGCACTCGTGTAAGGACACGAACCATCAGGC AGTTTCCCATTGGCAGTGAAAAGGAGTGTCCAGAATTTGAAGAAAAAGAACCCTGTTTGTCTCAAG GAGATGGAGTTGTCCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGCCGTGTGG ACCCTTTGCTCAGTCAGCAGGACAAGAGGCGCGGCAACCAGACGGCCCTCTGTGGAGGGGGCATCC AGACCCGAGAGGTGTACTGCGTGCAGGCCAACGAAAACCTCCTCTCACAATTAAGTACCCACAAGA ACAAAGAAGCCTCAAAGCCAATGGACTTAAAATTATGCACTGGACCTATCCCTAATACTACACAGC TGTGCCACATTCCTTGTCCAACTGAATGTGAAGTTTCACCTTGGTCAGCTTGGGGACCTTGTACTT ATGAAAACTGTAATGATCAGCAAGGAAAAAGGCTTCAAACTGAGGAAGCGGCGCATTACCAATG AGCCCACTGGAGGCTCTGGGGTAACCGGAAACTGCCCTCACTTACTGGAAGCCATTCCCTGTGAAG AGCCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGGAGT GTGGTCCAGGCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTGATGGAGAAGAAGTTGACAGAC AGCTGTGCAGAGATGCCATCTTCCCCATCCCTGTGGCCTGTGATGCCCCATGCCCGAAAGACTGTG TGCTCAGCACATGGTCTACGTGGTCCTCCTGCTCACACACCTGCTCAGGGAAAACGACAGAAGGGA AACAGATACGAGCACGATCCATTCTGGCCTATGCGGGTGAAGAAGGTGGAATTCGCTGTCCAAATA GCAGTGCTTTGCAAGAAGTACGAAGCTGTAATGAGCATCCTTGCACAGTGTACCACTGGCAAACTG GTCCCTGGGGCCAGTGCATTGAGGACACCTCAGTATCGTCCTTCAACACAACTACGACTTGGAATG GGGAGGCCTCCTGCTCTGTCGGCATGCAGACAAGAAAAGTCATCTGTGTGCGAGTCAATGTGGGCC AAGTGGGACCCAAAAAATGTCCTGAAAGCCTTCGACCTGAAACTGTAAGGCCTTGTCTGCTTCCTT GTAAGAAGGACTGTATTGTGACCCCATATAGTGACTGGACATCATGCCCCTCTTCGTGTAAAGAAG GGGACTCCAGTATCAGGAAGCAGTCTAGGCATCGGGTCATCATTCAGCTGCCAGCCAACGGGGGCC GAGACTGCACAGATCCCCTCTATGAAGAGAAGGCCTGTGAGGCACCTCAAGCGTGCCAAAGCTACA GGTGGAAGACTCACAAATGGCGCAGATGCCAATTAGTCCCTTGGAGCGTGCAACAAGACAGCCCTG GAGCACAGGAAGGCTGTGGGCCTGGGCGACAGGCAAGAGCCATTACTTGTCGCAAGCAAGATGGAG GACAGGCTGGAATCCATGAGTGCCTACAGTATGCAGGCCCTGTGCCAGCCCTTACCCAGGCCTGCC AGATCCCCTGCCAGGATGACTGTCAATTGACCAGCTGGTCCAAGTTTTCTTCATGCAATGGAGACT GTGGTGCAGTTAGGACCAGAAAGCGCACTCTTGTTGGAAAAAGTAAAAAGAAGGAAAAATGTAAAA ATTCCCATTTGTATCCCCTGATTGAGACTCAGTATTGTCCTTGTGACAAATATAATGCACAACCTG TGGGGAACTGGTCAGACTGTATTTTACCAGAGGGAAAAGTGGAAGTGTTGCTGGGAATGAAAGTAC AAGGAGACATCAAGGAATGCGGACAAGGATATCGTTACCAAGCAATGGCATGCTACGATCAAAATG GCAGGCTTGTGGAAACATCTAGATGTAACAGCCATGGTTACATTGAGGAGGCCTGCATCATCCCCT GCCCCTCAGACTGCAAGCTCAGTGAGTGGTCCAACTGGTCGCGCTGCAGCAAGTCCTGTGGGAGTG GTGTGAAGGTTCGTTCTAAATGGCTGCGTGAAAAACCATATAATGGAGGAAGGCCTTGCCCCAAAC TGGACCATGTCAACCAGGCACAGGTGTATGAGGTTGTCCCATGCCACAGTGACTGCAACCAGTACC TATGGGTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAGAACTGTGGAG AGGGCGTGCAAACCCGAAAAGTGAGATGCATGCAGAATACAGCAGATGGCCCTTCTGAACATGTAG AGGATTACCTCTGTGACCCAGAAGAGATGCCCCTGGGCTCTAGAGTGTGCAAATTACCATGCCCTG		

	AGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAATGTGTTTTGCCTTGCAATCAAAGCAGTT TCCGGCAAAGGTCAGCTGATCCCATCAGACAACCAGCTGATGAAGGAAGATCTTGCCCTAATGCTG TTGAGAAAGAACCCTGTAACCTGAACAAAAACTGCTACCACTATGATTATAATGTAACAGACTGGA GTACATGTCAGCTGAGTGAGAAGGCAGTTTGTGGAAATGGAATAAAAACAAGGATGTTGGATTGTG TTCGAAGTGATGGCAAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAGAACTGGC AGATGAACACGTCCTGCATGGTGGAATGCCCTGTGAACTGTCAGCTTTCTGATTGGTCTCCTTGGT CAGAATGTTCTCAAACATGTGGCCTCACAGGAAAAATGATCCGAAGACGAACAGTGACCCAGCCCT TTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCAGTCCAAACCCTGCCCAGTGAAGCCTT GTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCCCAGTGTGGAGAAGGGA CCAGAACAAGGAACATTTCTTGTGTAGTAAGTGATGGGTCAGCTGATGATTTCAGCAAAGTGGTGG ATGAGGAATTCTGTGCTGACATTGAACTCATTATAGATGGTAATAAAAATATGGTTCTGGAGGAAT CCTGCAGCCAGCCTTGCCCAGGTGACTGTTATTTGAAGGACTGGTCTTCCTGGAGCCTGTGTCAGC TGACCTGTGTGAATGGTGAGGATCTAGGCTTTGGTGGAATACAGGTCAGATCCAGACCGGTGATTA TACAAGAACTAGAGAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAATCATGTTATGATG GACAGTGCTATGAATATAAATGGATGGCCAGTGCTTGGAAGGGCTCTTCCCGAACAGTGTGGTGTC AAAGGTCAGATGGTATAAATGTAACAGGGGGCTGCTTGGTGATGAGCCAGCCTGATGCCGACAGGT CTTGTAACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAAACATGCCATTGTGAAG AAGGGTACACTGAAGTCATGTCTTCTAACAGCACCCTTGAGCAATGCACACTTATCCCCGTGGTG TATTACCCACCATGGAGGACAAAAGAGGAGATGTGAAAACCAGTCGGGCTGTACATCCAACCCAAC CCTCCAGTAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAGCCATTTGGGCCAGATGGGAGAC TAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTTGTGTTACTCATCTTTATTGTCTCCATGA TTTATCTAGCTTGCAAAAAGCCAAAGAAACCCCAAAGAAGGCAAAACAACCGACTGAAACCTTTAA CCTTAGCCTATGATGGAGATGCCGACATGTAACATATAACTTTTCCTGGCAACAACCA		
	ORF Start: ATG at 18		ORF Stop: TAA at 4782

	SEQ ID NO: 200	1588 aa	MW at 178042.1kD
NOV47m, CG51595-05 Protein Sequence	MGDECGPGGIQTRAVWCAHVEGWTTLHTNCKQAERPNNQQNCFKVCDWHKELYDWRLGPWNQCQPV ISKSLEKPLECIKGEEGIQVREIACIQKDKDIPAEDIICEYFEPKPLLEQACLIPCQQDCIVSEFS AWSECSKTCGSGLQHRTRHVVAPPQFGGSGCPNLTEFQVCQSSPCEAEELRYSLHVGPWSTCSMPH SRQVRQARRRGKNKEREKDRSKGVKDPEARELIKKKRNRNRQNRQENKYWDIQIGYQTREVMCINK TGKAADLSFCQQEKLPMTFQSCVITKECQVSEWSEWSPCSKTCHDMVSPAGTRVRTRTIRQFPIGS EKECPEFEEKEPCLSQGDGVVPCATYGWRTTEWTECRVDPLLSQQDKRRGNQTALCGGGIQTREVY CVQANENLLSQLSTHKNKEASKPMDLKLCTGPIPNTTQLCHIPCPTECEVSPWSAWGPCTYENCND QQGKKGFKLRKRRITNEPTGGSGVTGNCPHLLEAIPCEEPACYDWKAVRLGDCEPDNGKECGPGTQ VQEVVCINSDGEEVDRQLCRDAIFPIPVACDAPCPKDCVLSTWSTWSSCSHTCSGKTTEGKQIRAR SILAYAGEEGGIRCPNSSALQEVRSCNEHPCTVYHWQTGPWGQCIEDTSVSSFNTTTTWNGEASCS VGMQTRKVICVRVNVGQVGPKKCPESLRPETVRPCLLPCKKDCIVTPYSDWTSCPSSCKEGDSSIR KQSRHRVIIQLPANGGRDCTDPLYEEKACEAPQACQSYRWKTHKWRRCQLVPWSVQQDSPGAQEGC GPGRQARAITCRKQDGGQAGIHECLQYAGPVPALTQACQIPCQDDCQLTSWSKFSSCNGDCGAVRT RKRTLVGKSKKKEKCKNSHLYPLIETQYCPCDKYNAQPVGNWSDCILPEGKVEVLLGMKVQGDIKE CGQGYRYQAMACYDQNGRLVETSRCNSHGYIEEACIIPCPSDCKLSEWSNWSRCSKSCGSGVKVRS KWLREKPYNGGRPCPKLDHVNQAQVYEVVPCHSDCNQYLWVTEPWSICKVTFVNMRENCGEGVQTR KVRCMQNTADGPSEHVEDYLCDPEEMPLGSRVCKLPCPEDCVISEWGPWTQCVLPCNQSSFRQRSA DPIRQPADEGRSCPNAVEKEPCNLNKNCYHYDYNVTDWSTCQLSEKAVCGNGIKTRMLDCVRSDGK SVDLKYCEALGLEKNWQMNTSCMVECPVNCQLSDWSPWSECSQTCGLTGKMIRRRTVTQPFQGDGR PCPSLMDQSKPCPVKPCYRWQYGQWSPCQVQEAQCGEGTRTRNISCVVSDGSADDFSKVVDEEFCA DIELIIDGNKNMVLEESCSQPCPGDCYLKDWSSWSLCQLTCVNGEDLGFGGIQVRSRPVIIQELEN QHLCPEQMLETKSCYDGQCYEYKWMASAWKGSSRTVWCQRSDGINVTGGCLVMSQPDADRSCNPPC SQPHSYCSETKTCHCEEGYTEVMSSNSTLEQCTLIPVVVLPTMEDKRGDVKTSRAVHPTQPSSNPA GRGRTWFLQPFGPDGRLKTWVYGVAAGAFVLLIFIVSMIYLACKKPKKPQRRQNNRLKPLTLAYDG DADM		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 47B.

Table 47B. Comparison of NOV47a against NOV47b through NOV47m.		
Protein Sequence	**NOV47a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV47b	299..765 1..467	438/467 (93%) 439/467 (93%)

NOV47c	849..1419 1..571	562/571 (98%) 562/571 (98%)
NOV47d	1..1531 1..1531	1480/1531 (96%) 1480/1531 (96%)
NOV47e	1..1437 1..1437	1384/1437 (96%) 1384/1437 (96%)
NOV47f	1..1531 3..1533	1477/1531 (96%) 1479/1531 (96%)
NOV47g	1..1588 1..1588	1537/1588 (96%) 1537/1588 (96%)
NOV47h	102..632 3..533	489/531 (92%) 489/531 (92%)
NOV47i	102..632 3..533	488/531 (91%) 489/531 (91%)
NOV47j	1..1588 1..1588	1537/1588 (96%) 1537/1588 (96%)
NOV47k	851..1422 4..575	563/572 (98%) 563/572 (98%)
NOV47l	1..307 1..307	292/307 (95%) 292/307 (95%)
NOV47m	1..1588 1..1588	1537/1588 (96%) 1537/1588 (96%)

Further analysis of the NOV47a protein yielded the following properties shown in Table 47C.

Table 47C. Protein Sequence Properties NOV47a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.3500 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.2000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV47a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 47D.

Table 47D. Geneseq Results for NOV47a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV47a Residues/ Match	Identities/ Similarities for the Matched Region	Expect Value

		Residues		
AAB20155	Secreted protein SECP1 - Homo sapiens, 1588 aa. [WO200105971-A2, 25-JAN-2001]	1..1588 1..1588	1588/1588 (100%) 1588/1588 (100%)	0.0
AAM39295	Human polypeptide SEQ ID NO 2440 - Homo sapiens, 1588 aa. [WO200153312-A1, 26-JUL-2001]	1..1588 1..1588	1587/1588 (99%) 1588/1588 (99%)	0.0
AAM41081	Human polypeptide SEQ ID NO 6012 - Homo sapiens, 1551 aa. [WO200153312-A1, 26-JUL-2001]	48..1588 11..1551	1540/1541 (99%) 1540/1541 (99%)	0.0
AAB42496	Human ORFX ORF2260 polypeptide sequence SEQ ID NO:4520 - Homo sapiens, 617 aa. [WO200058473-A2, 05-OCT-2000]	1..614 6..617	605/614 (98%) 607/614 (98%)	0.0
AAM28984	Peptide #3021 encoded by probe for measuring placental gene expression - Homo sapiens, 277 aa. [WO200157272-A2, 09-AUG-2001]	1..271 6..276	271/271 (100%) 271/271 (100%)	e-169

In a BLAST search of public sequence datbases, the NOV47a protein was found to have homology to the proteins shown in the BLASTP data in Table 47E.

Table 47E. Public BLASTP Results for NOV47a

Protein Accession Number	Protein/Organism/Length	NOV47a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC32422	Sequence 1 from Patent WO0105971 - Homo sapiens (Human), 1588 aa.	1..1588 1..1588	1588/1588 (100%) 1588/1588 (100%)	0.0
BAA76804	KIAA0960 protein - Homo sapiens (Human), 1502 aa (fragment).	87..1588 1..1502	1502/1502 (100%) 1502/1502 (100%)	0.0
Q9UPZ6	KIAA0960 protein - Homo sapiens (Human), 1290 aa (fragment).	299..1588 1..1290	1290/1290 (100%) 1290/1290 (100%)	0.0

Q9C0I4	KIAA1679 protein - Homo sapiens (Human), 1536 aa (fragment).	22..1588 1..1536	790/1574 (50%) 1044/1574 (66%)	0.0
O43384	Hypothetical protein GS164B05.1 in chromosome 7 - Homo sapiens (Human), 446 aa (fragment).	954..1401 1..446	446/448 (99%) 446/448 (99%)	0.0

PFam analysis predicts that the NOV47a protein contains the domains shown in the Table 47F.

Table 47F. Domain Analysis of NOV47a

Pfam Domain	NOV47a Match Region	Identities/ Similarities for the Matched Region	Expect Value
tsp_1	129..177	20/54 (37%) 39/54 (72%)	1.8e-13
tsp_1	295..353	23/63 (37%) 45/63 (71%)	0.0015
tsp_1	447..504	21/62 (34%) 39/62 (63%)	0.1
tsp_1	569..625	19/60 (32%) 41/60 (68%)	0.0012
tsp_1	706..761	17/59 (29%) 40/59 (68%)	0.0014
tsp_1	841..889	17/57 (30%) 34/57 (60%)	0.022
tsp_1	970..1021	18/55 (33%) 40/55 (73%)	1.1e-05
tsp_1	1030..1093	14/69 (20%) 46/69 (67%)	0.79
tsp_1	1100..1150	15/55 (27%) 36/55 (65%)	0.039
tsp_1	1221..1271	20/55 (36%) 39/55 (71%)	6.9e-11
tsp_1	1349..1405	17/62 (27%) 34/62 (55%)	0.29

Example 48.

The NOV48 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 48A.

Table 48A. NOV48 Sequence Analysis			
	SEQ ID NO: 201	3149 bp	
NOV48a, CG57209-01 DNA Sequence	CTAAAGTTTTTTTCTTTGAATGACAGAACTACAGCATAATGCGTGGCTTCAACCTGCTCCTCTTCTG GGGATGTTGTGTTATGCACAGCTGGGAAGGGCACATAAGACCCACACGGAAACCAAACACAAAGGGT AATAACTGTAGAGACAGTACCTTGTGCCCAGCTTATGCCACCTGCACCAATACGGTGGACAGTTACT ATTGCACTTGCAAACAAGGCTTCCTGTCCAGCAATGGGCAAAATCACTTCAAGGATCCAGGAGTGCG ATGCAAAGATATTGATGAATGTTCTCAAAGCCCCCAGCCCTGTGGTCCTAACTCATCCTGCAAAAAC CTGTCAGGGAGGTACAAGTGCAGCTGTTTAGATGGTTTCTCTTCTCCCACTGGAAATGACTGGGTCC CAGGAAAGCCGGGCAATTTCTCCTGTACTGATATCAATGAGTGCCTCACCAGCAGGGTCTGCCCTGA GCATTCTGACTGTGTCAACTCCATGGGAAGCTACAGTTGCAGCTGTCAAGTTGGATTCATCTCTAGA AACTCCACCTGTGAAGACGTGAATGAATGTGCAGATCCAAGAGCTTGCCCAGAGCATGCAACTTGTA ATAACACTGTTGGAAACTACTCTTGTTTCTGCAACCCAGGATTTGAATCCAGCAGTGGCCACTTGAG TTGCCAGGGTCTCAAAGCATCGTGTGAAGATATTGATGAATGCACTGAAATGTGCCCCATCAATTCA ACATGCACCAACACTCCTGGGAGCTACTTTTGCACCTGCCACCCTGGCTTTGCACCAAGCAGTGGAC AGTTGAATTTCACAGACCAAGGAGTGGAATGTAGAGATATTGATGAGTGCCGCCAAGATCCATCAAC CTGTGGTCCTAATTCTATCTGCACCAATGCCCTGGGCTCCTACAGCTGTGGCTGCATTGTAGGCTTT CATCCCAATCCAGAAGGCTCCCAGAAAGATGGCAACTTCAGCTGCCAAAGGGTTCTCTTCAAATGTA AGGAAGATGTGATACCCGATAATAAGCAGATCCAGCAATGCCAAGAGGGAACCGCAGTGAAACCTGC ATATGTCTCCTTTTGTGCACAAATAAATAACATCTTCAGCGTTCTGGACAAAGTGTGTGAAAATAAA ACGACCGTAGTTTCTCTGAAGAATACAACTGAGAGCTTTGTCCCTGTGCTTAAACAAATATCCATGT GGACTAAATTCACCAAGGAAGAGACGTCCTCCCTGGCCACAGTCTTCCTGGAGAGTGTGGAAAGCAT GACACTGGCATCTTTTTGGAAACCCTCAGCAAATGTCACTCCGGCTGTTCGGGCGGAATACTTAGAC ATTGAGAGCAAAGTTATCAACAAAGAATGCAGTGAAGAGAATGTGACGTTGGACTTGGTAGCCAAGG GGGATAAGATGAAGATCGGGTGTTCCACAATTGAGGAATCTGAATCCACAGAGACCACTGGTGTGGC TTTTGTCTCCTTTGTGGGCATGGAATCGGTTTTAAATGAGCGCTTCTTCCAAGACCACCAGGCTCCC TTGACCACCTCTGAGATCAAGCTGAAGATGAATTCTCGAGTCGTTGGGGGCATAATGACTGGAGAGA AGAAAGACGGCTTCTCAGATCCAATCATCTACACTCTGGAGAACGTTCAGCCAAAGCAGAAGTTTGA GAGGCCCATCTGTGTTTCCTGGAGCACTGATGTGAAGGGTGGAAGATGGACATCCTTTGGCTGTGTG ATCCTGGAAGCTTCTGAGACATATACCATCTGCAGCTGTAATCAGATGGCAAATCTTGCCGTTATCA TGGCGTCTGGGGAGCTCACGATGGACTTTTCCTTGTACATCATTAGCCATGTAGGCATTATCATCTC CTTGGTGTGCCTCGTCTTGGCCATCGCCACCTTTCTGCTGTGTCGCTCCATCCGAAATCACAACACC TACCTCCACCTGCACCTCTGCGTGTGTCTCCTCTTGGCGAAGACTCTCTTCCTCGCCGGTATACACA AGACTGACAACAAGACGGGCTGCGCCATCATCGCGGGCTTCCTGCACTACCTTTTCCTTGCCTGCTT CTTCTGGATGCTGGTGGAGGCTGTGATACTGTTCTTGATGGTCAGAAACCTGAAGGTGGTGAATTAC TTCAGCTCTCGCAACATCAAGATGCTGCACATCTGTGCCTTTGGTTATGGGCTGCCGATGCTGGTGG TGGTGATCTCTGCCAGTGTGCAGCCACAGGGCTATGGAATGCATAATCGCTGCTGGCTGAATACAGA GACAGGGTTCATCTGGAGTTTCTTGGGGCCAGTTTGCACAGTTATAGTGATCAACTCCCTTCTCCTG ACCTGGACCTTGTGGATCCTGAGGCAGAGGCTTTCCAGTGTTAATGCCGAAGTCTCAACGCTAAAAG ACACCAGGTTACTGACCTTCAAGGCCTTTGCCCAGCTCTTCATCCTGGGCTGCTCCTGGGTGCTGGG CATTTTTCAGATTGGACCTGTGGCAGGTGTCATGGCTTACCTGTTCACCATCATCAACAGCCTGCAG GGGGCCTTCATCTTCCTCATCCACTGTCTGCTCAACGGCCAGGTACGAGAAGAATACAAGAGGTGGA TCACTGGGAAGACGAAGCCCAGCTCCCAGTCCCAGACCTCAAGGATCTTGCTGTCCTCCATGCCATC CGCTTCCAAGACGGGTTAAAGCCTTTCTTGCTTTCAAATATGCTATGGAGCCACAGTTGAGGACAGT AGTTTCCTGCAGGAGCCTACCCTGAAATCTCTTCTCAGCTTAACATGGAAATGAGGATCCCACCAGC CCCAGAACCCTCTGGGAAGAATGTTGGGGGCCGTCTTCCTGTGGTTGTATGCACTGATGAGAAATC AGACGTTTCTGCTCCAAACGACCATTTTATCTTCGTGCTCTGCAACTTCTTCAATTCCAGAGTTTCT GAGAACAGACCCAAATTCAATGGCATGACCAAGAACACCTGGCTACCATTTTGTTTTCTCCTGCCCT TGTTGGTGCATGGTTCTAAGCGTGCCCCTCCAGCGCCTATCATACGCCTGACACAGAGAACCTCTCA ATAAATGATTTGTCGCCTGTCTGACTGATTTACCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 39		ORF Stop: TAA at 2697

	SEQ ID NO: 202	886 aa	MW at 97679.1kD
NOV48a, CG57209-01 Protein Sequence	MRGFNLLLFWGCCVMHSWEGHIRPTRKPNTKGNNCRDSTLCPAYATCTNTVDSYYCTCKQGFLSSNG QNHFKDPGVRCKDIDECSQSPQPCGPNSSCKNLSGRYKCSCLDGFSSPTGNDWVPGKPGNFSCTDIN ECLTSRVCPEHSDCVNSMGSYSCSCQVGFISRNSTCEDVNECADPRACPEHATCNNTVGNYSCFCNP GFESSSGHLSCQGLKASCEDIDECTEMCPINSTCTNTPGSYFCTCHPGFAPSSGQLNFTDQGVECRD IDECRQDPSTCGPNSICTNALGSYSCGCIVGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNKQIQQ CQEGTAVKPAYVSFCAQINNIFSVLDKVCENKTTVVSLKNTTESFVPVLKQISMWTKFTKEETSSLA TVFLESVESMTLASFWKPSANVTPAVRAEYLDIESKVINKECSEENVTLDLVAKGDKMKIGCSTIEE SESTETTGVAFVSFVGMESVLNERFFQDHQAPLTTSEIKLKMNSRVVGGIMTGEKKDGFSDPIIYTL ENVOPKOKFERPICVSWSTDVKGGRWTSFGCVILEASETYTICSCNOMANLAVIMASGELTMDFSLY		

	IIISHVGIIISLVCLVLAIATFLLCRSIRNHNTYLHLHLCVCLLLAKTLFLAGIHKTDNKTGCAIIAG FLHYLFLACFFWMLVEAVILFLMVRNLKVVNYFSSRNIKMLHICAFGYGLPMLVVVISASVQPQGYG MHNRCWLNTETGFIWSFLGPVCTVIVINSLLLTWTLWILRQRLSSVNAEVSTLKDTRLLTFKAFAQL FILGCSWVLGIFQIGPVAGVMAYLFTIINSLQGAFIFLIHCLLNGQVREEYKRWITGKTKPSSQSQT SRILLSSMPSASKTG

	SEQ ID NO: 203	1764 bp	
NOV48b, CG57209-03 DNA Sequence	AGATCTTGGGAAGGGCACATAAGACCCACACGGAAACCAAACACAAAGGGTAATAACTGTAGAGACA GTACCTTGTGCCCAGCTTATGCCACCTGCACCAATACAGTGGACAGTTACTATTGCACTTGCAAACA AGGCTTCCTGTCCAGCAATGGGCAAAATCACTTCAAGGATCCAGGAGTGCGATGCAAAGATATTGAT GAATGTTCTCAAAGCCCCCAGCCCTGTGGTCCTAACTCATCCTGCAAAAACCTGTCAGGGAGGTACA AGTGCAGCTGTTTAGATGGTTTCTCTTCTCCCACTGGAAATGACTGGGTCCCAGGAAAGCCGGGCAA TTTCTCCTGTACTGATATCAATGAGTGCCTCACCAGCAGGGTCTGCCCTGAGCATTCTGACTGTGTC AACTCCATGGGAAGCTACAGTTGCAGCTGTCAAGTTGGATTCATCTCTAGAAACTCCACCTGTGGAG ACGTGAATGAATGTGCAGATCCAAGAGCTTGCCCAGAGCATGCAACTTGTAATAACACTGTTGGAAA CTACTCTTGTTTCTGCAACCCAGGATTTGAATCCAGCAGTGGCCACTTGAGTTTCCAGGGTCTCAAA GCATCGTGTGAAGATATTGATGAATGCACTGAAATGTGCCCCATCAATTCAACATGCACCAACACTC CTGGGAGCTACTTTTGCACCTGCCACCCTGGCTTTGCACCAAGCAATGGACAGTTGAATTTCACAGA CCAAGGAGTGGAATGTAGAGATATTGATGAGTGCCGCCAAGATCCATCAACCTGTGGTCCTAATTCT ATCTGCACCAATGCCCTGGGCTCCTACAGCTGTGGCTGCATTGTAGGCTTTCATCCCAATCCAGAAG GCTCCCAGAAAGATGGCAACTTCAGCTGTCAAAGGGTTCTCTTCAAATGTAAGGAAGATGTGATACC CGATAATAAGCAGATCCAGCAATGCCAAGAGGGAACCGCAGTGAAACCTGCATATGTCTCCTTTTGT GCACAAATAAATAACATCTTCAGCGTTCTGGACAAAGTGTGTGAAAATAAAACGACCGTAGTTTCTC TGAAGAATACAACTGAGAGCTTTGTCCCTGTGCTTAAACAAATATCCACGTGGACTAAATTCACCAA GGAAGAGACGTCCTCCCTGGCCACAGTCTTCCTGGAGAGTGTGGAAAGCATGACACTGGCATCTTTT TGGAAACCCTCAGCAAATGTCACTCCGGCTGTTCGGACGGAATACTTAGACATTGAGAGCAAAGTTA TCAACAAAGAATGCAGTGAAGAGAATGTGACGTTGGACTTGGTAGCCAAGGGGGATAAGATGAAGAT CGGGTGTTCCACAATTGAGGAATCTGAATCCACAGAGACCACTGGTGTGGCTTTTGTCTCCTTTGTG GGCATGGAATCGGTTTTAAATGAGCGCTTCTTCCAAGACCACCAGGCTCCCTTGACCACCTCTGAGA TCAAGCTGAAGATGAATTCTCGAGTCGTTGGGGGCATAATGACTGGAGAGAAGAAAGACGGCTTCTC AGATCCAATCATCTACACTCTGGAGAACGTTCAGCCAAAGCAGAAGTTTGAGAGGCCCATCTGTGTT TCCTGGAGCACTGATGTGAAGGGTGGAAGATGGACATCCTTTGGCTGTGTGATCCTGGAAGCTTCTG AGACATATACCATCTGCAGCTGTAATCAGATGGCAAATCTTGCCGTTATCATGGCGTCTGGGGAGCT CACGGTCGACAAGGGCGAATTT		
	ORF Start: at 7		ORF Stop: at 1747

	SEQ ID NO: 204	580 aa	MW at 63248.2kD
NOV48b, CG57209-03 Protein Sequence	WEGHIRPTRKPNTKGNNCRDSTLCPAYATCTNTVDSYYCTCKQGFLSSNGQNHFKDPGVRCKDIDEC SQSPQPCGPNSSCKNLSGRYKCSCLDGFSSPTGNDWVPGKPGNFSCTDINECLTSRVCPEHSDCVNS MGSYSCSCQVGFISRNSTCGDVNECADPRACPEHATCNNTVGNYSCFCNPGFESSSGHLSFQGLKAS CEDIDECTEMCPINSTCTNTPGSYFCTCHPGFAPSNGQLNFTDQGVECRDIDECRQDPSTCGPNSIC TNALGSYSCGCIVGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNKQIQQCQEGTAVKPAYVSFCAQ INNIFSVLDKVCENKTTVVSLKNTTESFVPVLKQISTWTKFTKEETSSLATVFLESVESMTLASFWK PSANVTPAVRTEYLDIESKVINKECSEENVTLDLVAKGDKMKIGCSTIEESESTETTGVAFVSFVGM ESVLNERFFQDHQAPLTTSEIKLKMNSRVVGGIMTGEKKDGFSDPIIYTLENVQPKQKFERPICVSW STDVKGGRWTSFGCVILEASETYTICSCNQMANLAVIMASGELT		

	SEQ ID NO: 205	1740 bp	
NOV48c, CG57209-02 DNA Sequence	TGGGAAGGGCACATAAGACCCACACGGAAACCAAACACAAAGGGTAATAACTGTAGAGACAGTACCT TGTGCCCAGCTTATGCCACCTGCACCAATACAGTGGACAGTTACTATTGCACTTGCAAACAAGGCTT CCTGTCCAGCAATGGGCAAAATCACTTCAAGGATCCAGGAGTGCGATGCAAAGATATTGATGAATGT TCTCAAAGCCCCCAGCCCTGTGGTCCTAACTCATCCTGCAAAAACCTGTCAGGGAGGTACAAGTGCA GCTGTTTAGATGGTTTCTCTTCTCCCACTGGAAATGACTGGGTCCCAGGAAAGCCGGGCAATTTCTC CTGTACTGATATCAATGAGTGCCTCACCAGCAGGGTCTGCCCTGAGCATTCTGACTGTGTCAACTCC ATGGGAAGCTACAGTTGCAGCTGTCAAGTTGGATTCATCTCTAGAAACTCCACCTGTGGAGACGTGA ATGAATGTGCAGATCCAAGAGCTTGCCCAGAGCATGCAACTTGTAATAACACTGTTGGAAACTACTC TTGTTTCTGCAACCCAGGATTTGAATCCAGCAGTGGCCACTTGAGTTTCCAGGGTCTCAAAGCATCG TGTGAAGATATTGATGAATGCACTGAAATGTGCCCCATCAATTCAACATGCACCAACACTCCTGGGA GCTACTTTTGCACCTGCCACCCTGGCTTTGCACCAAGCAATGGACAGTTGAATTTCACAGACCAAGG AGTGGAATGTAGAGATATTGATGAGTGCCGCCAAGATCCATCAACCTGTGGTCCTAATTCTATCTGC ACCAATGCCCTGGGCTCCTACAGCTGTGGCTGCATTGTAGGCTTTCATCCCAATCCAGAAGGCTCCC AGAAAGATGGCAACTTCAGCTGTCAAAGGGTTCTCTTCAAATGTAAGGAAGATGTGATACCCGATAA TAAGCAGATCCAGCAATGCCAAGAGGGAACCGCAGTGAAACCTGCATATGTCTCCTTTTGTGCACAA ATAAATAACATCTTCAGCGTTCTGGACAAAGTGTGTGAAAATAAAACGACCGTAGTTTCTCTGAAGA ATACAACTGAGAGCTTTGTCCCTGTGCTTAAACAAATATCCACGTGGACTAAATTCACCAAGGAAGA		

	GACGTCCTCCCTGGCCACAGTCTTCCTGGAGAGTGTGGAAAGCATGACACTGGCATCTTTTTGGAAA CCCTCAGCAAATGTCACTCCGGCTGTTCGGACGGAATACTTAGACATTGAGAGCAAAGTTATCAACA AAGAATGCAGTGAAGAGAATGTGACGTTGGACTTGGTAGCCAAGGGGGATAAGATGAAGATCGGGTG TTCCACAATTGAGGAATCTGAATCCACAGAGACCACTGGTGTGGCTTTTGTCTCCTTTGTGGGCATG GAATCGGTTTTAAATGAGCGCTTCTTCCAAGACCACCAGGCTCCCTTGACCACCTCTGAGATCAAGC TGAAGATGAATTCTCGAGTCGTTGGGGGCATAATGACTGGAGAGAAGAAAGACGGCTTCTCAGATCC AATCATCTACACTCTGGAGAACGTTCAGCCAAAGCAGAAGTTTGAGAGGCCCATCTGTGTTTCCTGG AGCACTGATGTGAAGGGTGGAAGATGGACATCCTTTGGCTGTGTGATCCTGGAAGCTTCTGAGACAT ATACCATCTGCAGCTGTAATCAGATGGCAAATCTTGCCGTTATCATGGCGTCTGGGGAGCTCACG	
	ORF Start: at 1	ORF Stop: end of sequence

	SEQ ID NO: 206	580 aa	MW at 63248.2kD
NOV48c, CG57209-02 Protein Sequence	WEGHIRPTRKPNTKGNNCRDSTLCPAYATCTNTVDSYYCTCKQGFLSSNGQNHFKDPGVRCKDIDEC SQSPQPCGPNSSCKNLSGRYKCSCLDGFSSPTGNDWVPGKPGNFSCTDINECLTSRVCPEHSDCVNS MGSYSCSCQVGFISRNSTCGDVNECADPRACPEHATCNNTVGNYSCFCNPGFESSSGHLSFQGLKAS CEDIDECTEMCPINSTCTNTPGSYFCTCHPGFAPSNGQLNFTDQGVECRDIDECRQDPSTCGPNSIC TNALGSYSCGCIVGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNKQIQQCQEGTAVKPAYVSFCAQ INNIFSVLDKVCENKTTVVSLKNTTESFVPVLKQISTWTKFTKEETSSLATVFLESVESMTLASFWK PSANVTPAVRTEYLDIESKVINKECSEENVTLDLVAKGDKMKIGCSTIEESESTETTGVAFVSFVGM ESVLNERFFQDHQAPLTTSEIKLKMNSRVVGGIMTGEKKDGFSDPIIYTLENVQPKQKFERPICVSW STDVKGGRWTSFGCVILEASETYTICSCNQMANLAVIMASGELT		

	SEQ ID NO: 207	2851 bp	
NOV48d, CG57209-04 DNA Sequence	GCTCCTCTTCTG**GGG**TGTTGTGTTATGCACAGCTGGGAAGGGCACATAAGACCCACACGGAAACCAA ACACAAAGGGTAATAACTGTAGAGACAGTACCTTGTGCCCAGCTTATGCCACCTGCACCAATACAGT GGACAGTTACTATTGCGCTTGCAAACAAGGCTTCCTGTCCAGCAATGGGCAAAATCACTTCAAGGAT CCAGGAGTGCGATGCAAAGATATTGATGAATGTTCTCAAAGCCCCCAGCCCTGTGGTCCTAACTCAT CCTGCAAAAACCTGTCAGGGAGGTACAAGTGCAGCTGTTTAGATGGTTTCTCTTCTCCCACTGGAAA TGACTGGGTCCCAGGAAAGCCGGGCAATTTCTCCTGTACTGATATCAATGAGTGCCTCACCAGCAGC GTCTGCCCTGAGCATTCTGACTGTGTCAACTCCATGGGAAGCTACAGTTGTAGCTGTCAAGTTGGAT TCATCTCTAGAAACTCCACCTGTGAAGACGTGGATGAATGTGCAGATCCAAGAGCTTGCCCAGAGCA TGCAACTTGTAATAACACTGTTGGAAACTACTCTTGTTTCTGCAACCCAGGATTTGAATCCAGCAGT GGCCACTTGAGTTTCCAGGGTCTCAAAGCATCGTGTGAAGATATTGATGAATGCACTGAAATGTGCC CCATCAATTCAACATGCACCAACACTCCTGGGAGCTACTTTTGCACCTGCCACCCTGGCTTTGCACC AAGCAATGGACAGTTGAATTTCACAGACCAAGGAGTGGAATGTAGAGATATTGATGAGTGCCGCCAA GATCCATCAACCTGTGGTCCTAATTCTATCTGCACCAATGCCCTGGGCTCCTGCAGCTGTGGCTGCA TTGCAGGCTTTCATCCCAATCCAGAAGGCTCCCAGAAAGATGGCAACTTCAGCTGCCAAAGGGTTCT CTTCAAATGTAAGGAAGATGTGATACCCGATAATAAGCAGATCCAGCAATGCCAAGAGGGAACCGCA GTGAAACCTGCATATGTCTCCTTTTGTGCACAAATAAATAACATCTTCAGCGTTCTGGACAAAGTGT GTGAAAATAAAACGACCGTAGTTTCTCTGAAGAATACAACTGAGAGCTTTGTCCCTGTGCTTAAACA AATATCCACGTGGACTAAATTCACCAAGGAAGAGACGTCCTCCCTGGCCACAGTCTTCCTGGAGAGT GTGGAAAGCATGACACTGGCATCTTTTTGGAAACCCTCAGCAAATGTCACTCCGGCTGTTCGGACGG AATACTTAGACATTGAGAGCAAAGTTATCAACAAAGAATGCAGTGAAGAGAATGTGACGTTGGACTT GGTAGCCAAGGGGGATAAGATGAAGATCGGGTGTTCCACAATTGAGGAATCTGAATCCACAGAGACC ACTGGTGTGGCTTTTGTCTCCTTTGTGGGCATGGAATCGGTTTTAAATGAGCGCTTCTTCCAAGACC ACCAGGCTCCCTTGACCACCTCTGAGATCAAGCTGAAGATGAATTCTCGAGTCGTTGGGGGCATAAT GACTGGAGAGAAGAAAGACGGCTTCTCAGATCCAATTATCTACACTCTGGAGAACGTTCAGCCAAAG CAGAAGTTTGAGAGGCCCATCTGTGTTTCCTGGAGCACTGATGTGAAGGGTGGAAGATGGACATCCT TTGGCTGTGTGATCCTGGAAGCTTCTGAGACATATACCATCTGCAGCTGTAATCAGATGGCAAATCT TGCCGTTATCATGGCGTCTGGGGAGCTCACGATGGCTGCGCCATCATCGCGGGCTTCCTGCACTAC CTTTTCCTTGCCTGCTTCTTCTGGATGCTGGTGGAGGCTGTGATACTGTTCTTGATGGTCAGAAACC TGAAGGTGGTGAATTACTTCAGCTCTCGCAACATCAAGATGCTGCACATCTGTGCCTTTGGTTATGG GCTGCCGATGCTGGTGGTGGTGATCTCTGCCAGTGTGCAGCCACAGGGCTATGGAATGCATAATCGC TGCTGGCTGAATACAGAGACAGGGTTCATCTGGAGTTTCTTGGGGCCAGTTTGCACAGTTATAGTGA TCAACTCCCTTCTCCTGACCTGGACCTTGTGGATCCTGAGGCAGAGGCTTTCCAGTGTTAATGCCGA AGTCTCAACGCTAAAAGACACCAGGTTACTGACCTTCAAGGCCTTTGCCCAGCTCTTCATCCTGGGC TGCTCCTGGGTGCTGGGCATTTTTCAGATTGGACCTGTGGCAGGTGTCATGGCTTACCTGTTCACCA TCATCAACAGCCTGCAGGGGGCCTTCATCTTCCTCATCCACTGTCTGCTCAACGGCCAGGTACGAGA AGAATACAAGAGGTGGATCACTGGGAAGACGAAGCCCAGCTCCCAGTCCCAGACCTCAAGGATCTTG CTGTCCTCCATGCCATCCGCTTCCAAGACGGT**TAA**AGTCCTTTCTTGCTTTCAAATATGCTATGGA GCCACAGTTGAGGACAGTAGTTTCCTGCAGGAGCCTACCCTGAAATCTCTTCTCAGCTTAACATGGA AATGAGGATCCCACCAGCCCCAGAACCCTCTGGGGAAGAATGTTGGGGGCCGTCTTCCTGTGGTTGT ATGCACTGATGAGAAATCAGGCGTTTCTGCTCCAAACGACCATTTTATCTTCGTGCTCTGCAACTTC TTCAATTCCAGAGTTTCTGAGAACAGACCCAAATTCAATGGCATGACCAAGAACACCTGGCTACCAT TTTGTTTTCTCCTGCCCTTGTTGGTGCATGGTTCTAAGCGTGCCCCTCCAGCGCCTATCATACGCCT GACACAGAGAACCTCTCAATAAATGATTTGTCGCCTG		
	ORF Start: at 13		ORF Stop: TAA at 2446

	SEQ ID NO: 208	811 aa	MW at 89011.6kD
NOV48d, CG57209-04 Protein Sequence	GCCVMHSWEGHIRPTRKPNTKGNNCRDSTLCPAYATCTNTVDSYYCACKQGFLSSNGQNHFKDPGVR CKDIDECSQSPQPCGPNSSCKNLSGRYKCSCLDGFSSPTGNDWVPGKPGNFSCTDINECLTSSVCPE HSDCVNSMGSYSCSCQVGFISRNSTCEDVDECADPRACPEHATCNNTVGNYSCFCNPGFESSSGHLS FQGLKASCEDIDECTEMCPINSTCTNTPGSYFCTCHPGFAPSNGQLNFTDQGVECRDIDECRQDPST CGPNSICTNALGSCSCGCIAGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNKQIQQCQEGTAVKPA YVSFCAQINNIFSVLDKVCENKTTVVSLKNTTESFVPVLKQISTWTKFTKEETSSLATVFLESVESM TLASFWKPSANVTPAVRTEYLDIESKVINKECSEENVTLDLVAKGDKMKIGCSTIEESESTETTGVA FVSFVGMESVLNERFFQDHQAPLTTSEIKLKMNSRVVGGIMTGEKKDGFSDPIIYTLENVQPKQKFE RPICVSWSTDVKGGRWTSFGCVILEASETYTICSCNQMANLAVIMASGELTMGCAIIAGFLHYLFLA CFFWMLVEAVILFLMVRNLKVVNYFSSRNIKMLHICAFGYGLPMLVVVISASVQPQGYGMHNRCWLN TETGFIWSFLGPVCTVIVINSLLLTWTLWILRQRLSSVNAEVSTLKDTRLLTFKAFAQLFILGCSWV LGIFQIGPVAGVMAYLFTIINSLQGAFIFLIHCLLNGQVREEYKRWITGKTKPSSQSQTSRILLSSM PSASKTG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 48B.

Table 48B. Comparison of NOV48a against NOV48b through NOV48d.		
Protein Sequence	**NOV48a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV48b	18..597 1..580	563/580 (97%) 564/580 (97%)
NOV48c	18..597 1..580	563/580 (97%) 564/580 (97%)
NOV48d	11..886 1..811	783/876 (89%) 788/876 (89%)

Further analysis of the NOV48a protein yielded the following properties shown in Table 48C.

Table 48C. Protein Sequence Properties NOV48a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 18 and 19

A search of the NOV48a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 48D.

Table 48D. Geneseq Results for NOV48a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB71869	Human EMR1 seven transmembrane domain - Homo sapiens, 886 aa. [WO200109328-A1, 08-FEB-2001]	1..886 1..886	886/886 (100%) 886/886 (100%)	0.0
AAB01249	Human EMR1 hormone receptor - Homo sapiens, 880 aa. [WO200034473-A2, 15-JUN-2000]	1..886 1..880	880/886 (99%) 880/886 (99%)	0.0
AAE17043	Human CD 97 protein - Homo sapiens, 835 aa. [WO200202602-A2, 10-JAN-2002]	74..872 16..817	272/853 (31%) 422/853 (48%)	e-122
AAB15728	Human CD97 protein - Homo sapiens, 835 aa. [WO200052039-A2, 08-SEP-2000]	74..872 16..817	272/853 (31%) 422/853 (48%)	e-122
AAY41090	Human CD97 protein - Homo sapiens, 835 aa. [WO9945111-A1, 10-SEP-1999]	74..872 16..817	272/853 (31%) 422/853 (48%)	e-122

In a BLAST search of public sequence datbases, the NOV48a protein was found to have homology to the proteins shown in the BLASTP data in Table 48E.

Table 48E. Public BLASTP Results for NOV48a				
Protein Accession Number	**Protein/Organism/Length**	**NOV48a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q14246	Cell surface glycoprotein EMR1 precursor (EMR1 hormone receptor) - Homo sapiens (Human), 886 aa.	1..886 1..886	886/886 (100%) 886/886 (100%)	0.0
BAC06133	Seven transmembrane helix receptor - Homo sapiens (Human), 929 aa.	11..885 29..905	866/877 (98%) 868/877 (98%)	0.0

Q61549	Cell surface glycoprotein EMR1 precursor (EMR1 hormone receptor) (Cell surface glycoprotein F4/80) - Mus musculus (Mouse), 931 aa.	1..886 1..931	606/937 (64%) 709/937 (74%)	0.0
BAC06178	Seven transmembrane helix receptor - Homo sapiens (Human), 837 aa.	74..872 18..819	272/853 (31%) 422/853 (48%)	e-121
O00718	CD97 - Homo sapiens (Human), 835 aa.	74..872 16..817	272/853 (31%) 422/853 (48%)	e-121

PFam analysis predicts that the NOV48a protein contains the domains shown in the Table 48F.

Table 48F. Domain Analysis of NOV48a			
Pfam Domain	**NOV48a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
EGF	35..70	13/47 (28%) 26/47 (55%)	0.29
TILa	34..89	16/58 (28%) 36/58 (62%)	0.42
EGF	176..212	15/47 (32%) 25/47 (53%)	0.0038
EGF	225..255	13/47 (28%) 23/47 (49%)	0.29
GPS	546..596	19/54 (35%) 46/54 (85%)	1.5e-18
7tm_2	599..851	96/276 (35%) 228/276 (83%)	9.2e-104

Example 49.

The NOV49 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 49A.

Table 49A. NOV49 Sequence Analysis			
	SEQ ID NO: 209	5184 bp	
NOV49a, CG57292-01 DNA Sequence	CCCCGCAGGGGAAGGCGGGTCCTGGCGGCCAGCGCGCGGTCCGCGCCCACCCTAGCCGACGGGGCCG GCAGAGCGCGCGGCGTCGGTGCCCTTGACCATGGCGGCGGCTGCGCTTCTGCTGGGGCTGGCGCTGC TGGCACCGCGGGCGGCCGGCGCGGGCATGGGCGCGTGCTATGACGGCGCAGGGCGCCCGCAGCGCTG CCTGCCGGTGTTCGAGAACGCGGCGTTTGGGCGGCTCGCCCAGGCCTCGCACACGTGCGGCAGCCCG CCCGAGGACTTCTGTCCCCACGTGGGCGCCGCGGGCGCGGGGGCTCATTGCCAGCGCTGCGACGCCG		

CCGACCCCCAGCGCCACCACAACGCCTCCTACCTCACCGACTTCCACAGCCAGGACGAGAGCACCTG
GTGGCAGAGCCCGTCCATGGCCTTCGGCGTGCAGTACCCCACCTCGGTCAACATCACCCTCCGCCTA
GGGAAGGCTTATGAGATCACGTATGTGAGGCTGAAGTTCCACACCAGTCGCCCTGAGAGCTTTGCCA
TCTACAAGCGCAGCCGCGCCGACGGCCCATGGGAGCCCTACCAGTTCTACAGCGCCTCCTGCCAGAA
GACCTACGGCCGGCCCGAGGGCCAGTACCTGCGCCCCGGCGAGGACGAGCGCGTGGCCTTCTGCACC
TCTGAGTTCAGCGACATCTCCCCGCTGAGTGGCGGCAACGTGGCCTTCTCCACCCTGGAGGGCCGGC
CCAGCGCCTACAACTTCGAGGAGAGCCCTGGGCTGCAGGAGTGGGTCACCAGCACCGAACTCCTCAT
CTCTCTAGACCGGCTCAACACGTTTGGGGACGACATCTTCAAGGACCCCAAGGTGCTCCAGTCCTAC
TATTATGCCGTGTCCGACTTCTCTGTGGGCGGCAGGTGCAAGTGCAACGGGCATGCCAGCGAGTGCG
GCCCCGACGTGGCAGGCCAGTTGGCCTGCCGGTGCCAGCACAACACCACCGGCACAGACTGTGAGCG
CTGCCTGCCCTTCTTCCAGGACCGCCCGTGGGCCCGGGGCACCGCCGAGGCTGCCCACGAGTGTCTG
CCCTGCAACTGCAGTGGCCGCTCCGAGGAATGCACGTTTGATCGGGAGCTCTTCCGCAGCACAGGCC
ACGGCGGGCGCTGTCACCACTGCCGTGACCACACAGCTGGGCCACACTGTGAGCGCTGTCAGGAGAA
TTTCTATCACTGGGACCCGCGGATGCCATGCCAGCCCTGTGACTGCCAGTCGGCAGGCTCCCTACAC
CTCCAGTGCGATGACACAGGCACCTGCGCCTGCAAGCCCACAGTGACTGGCTGGAAGTGTGACCGCT
GTCTGCCCGGGTTCCACTCGCTCAGTGAGGGAGGCTGCAGACCCTGCACTTGCAATCCCGCTGGCAG
CCTGGACACCTGTGACCCCCGCAGTGGGCGCTGCCCCTGCAAAGAGAATGTGGAAGGCAACCTATGT
GACAGATGTCGCCCGGGGACCTTTAACCTGCAGCCCCACAATCCAGCTGGCTGCAGCAGCTGTTTCT
GCTATGGCCACTCCAAGGTGTGCGCGTCCACTGCCCAGTTCCAGGTGCATCACATCCTCAGCGATTT
CCACCAGGGAGCCGAAGGCTGGTGGGCCAGAAGTGTGGGGGGCTCTGAGCACTCCCCACAATGGAGC
CCAAATGGGGTCCTCCTGAGCCCAGAAGACGAGGAGGAGCTCACAGCACCAGGGAAGTTCCTGGGAG
ACCAGCGGTTCAGCTATGGGCAGCCCCTCATACTGACCTTCCGGGTGCCCCCCGGGGACTCCCCACT
CCCTGTACAGCTGAGGCTGGAAGGGACAGGCTTGGCCCTGTCCCTGAGGCACTCTAGCCTGTCTGGC
CCCCAGGATGCCAGGGCATCCCAGGGAGGTAGAGCTCAGGTTCCACTGCAGGAGACCTCCGAGGACG
TGGCCCCTCCACTGCCCCCCTTCCACTTCCAGCGGCTCCTCGCCAACCTGACCAGCCTCCGCCTCCG
CGTCAGTCCCGGCCCCAGCCCTGCCGGTCCAGTGTTCCTGACTGAGGTCCGGCTCACATCCGCCCGG
CCAGGGCTTTCCCCGCCAGCCTCCTGGGTGGAGATTTGTTCATGTCCCACTGGCTACACGGGCCAGT
TCTGTGAATCCTGTGCTCCGGGATACAAGAGGGAGATGCCACAGGGGGGTCCCTATGCCAGCTGTGT
CCCCTGCACCTGTAACCAGCATGGCACCTGTGACCCCAACACAGGGATCTGTGTCTGCAGCCACCAT
ACCGAGGGCCCATCCTGTGAACGCTGTTTGCCAGGTTTCTATGGCAACCCCTTTCGCGGGCCAAGCCG
ACGACTGCCAGCCCTGTCCCTGCCCTGGCCAGTCGGCCTGTACGACCATCCCAGAGAGCGGGGAGGT
GGTGTGTACCCACTGCCCCCCGGGCCAGAGAGGGCGGCGCTGTGAGGTCTGTGATGATGGCTTTTTT
GGGGACCCGCTGGGGCTCTTTGGGCACCCCCAGCCCTGCCACCAGTGCCAGTGTAGCGGGAACGTGG
ACCCCAATGCCGTGGGCAACTGTGACCCCCTGTCTGGCCACTGCCTGCGCTGCCTGCACAACACCAC
GGGTGACCACTGTGAGCACTGTCAGGAAGGCTTCTACGGGAGCGCCCTGGCCCCTCGACCCGCAGAC
AAATGCATGCCTTGCAGCTGTCACCCACAGGGCTCGGTCAGTGAGCAGATGCCCTGCGACCCAGTGA
CAGGCCAATGCTCCTGCCTGCCTCATGTGACTGCACGGGACTGCAGCCGCTGCTACCCTGGCTTCTT
CGACCTCCAGCCTGGGAGGGGCTGCCGGAGCTGCAAGTGTCACCCACTGGGCTCCCAGGAGGACCAG
TGCCATCCCAAGACTGGACAGTGCACCTGCCGCCCAGGTGTCACAGGCCAGGCCTGTGACAGGTGCC
AGCTGGGTTTCTTCGGCTCCTCAATCAAGGGCTGCCGGGCCTGCAGGTGCTCCCCACTGGGCGCTGC
CTCGGCCCAGTGCCACTATAACGGCACATGCGTGTGCAGGCCTGGCTTCGAGGGCTACAAATGTGAC
CGCTGCCACTACAACTTCTTCCTCACGGCAGACGGCACACACTGCCAGCAATGTCCGTCCTGCTACG
CCCTGGTGAAGGAGGAGACAGCCAAGCTGAAGGCCAGACTGACTTTGACGGAGGGGTGGCTCCAAGG
GTCCGACTGTGGCAGTCCCTGGGACCACTAGACATTCTGCTGGAGAGGCCCCAAGGGGGACGTC
TACCAGGGCCATCACCTGCTTCCAGGGGCTCGGGAAGCCTTCCTGGAGCAGATGATGGGCCTCGAGG
GTGCTGTCAAGGCCGCCCGGGAGCAGCTGCAGAGGCTGAACAAGGGTGCCCGCTGTGCCCAGGCCGG
ATCCCAGAAGACCTGCACCCAGCTGGCAGACCTGGAGGCAGTGCTGGAGTCCTCGGAAGAGGAGATT
CTGCATGCAGCTGCCATTCTCGCGTCTCTGGAGATTCCTCAGGAAGGTCCCAGTCAGCCGACCAAAT
GGAGCCACCTGGCCATAGAGGCCCGTGCCCTCGCCAGGAGCCACAGAGACACCGCCACCAAGATCGC
AGCCACTGCTTGGAGGGCCCTGCTCGCCTCCAACACCAGCTACGCGCTTCTCTGGAATCTGCTGGAG
GGAAGGGTGGCCCTAGAGACCCAGCGGGACCTGGAGGACAGGTACCAGGAGGTCCAGGCGGCCCAGA
AAGCACTGAGGACGGCTGTGGCAGAGGTGCTGCCTGAAGCGGAAAGCGTGTTGGCCACCGTGCAGCA
AGTTGGCGCAGATACAGCCCCGTACCTGGCCTTGCTGGCTTCCCCGGGAGCTCTGCCTCAGAAGTCC
CGGGCTGAAGACCTGGGCCTGAAGGCGAAGGCCCTGGAGAAGACAGTTGCATCATGGCAGCACATGG
CCACTGAGGCTGCCCGAACCCTCCAGACTGCTGCCCAGGCGACGCTACGGCAAACAGAACCCCTCAC
AATGGCGCGATCTCGGCTCACTGCAACCTTTGCCTCCCAGCTGCACCAGGGGGCCAGAGCCGCCCTG
ACCCAGGCTTCCTCATCTGTCCAGGCTGCGACAGTGACTGTCATGGGAGCCAGGACTCTGCTGGCTG
ATCTGGAAGGAATGAAGCTGCAGTTTCCCCGGCCCAAGGACCAGGCGGCATTGCAGAGGAAGGCAGA
CTCCGTCAGTGACAGACTCCTTGCAGACACGAGAAAGAAGACCAAGCAGGCGGAGAGGATGCTGGGA
AACGCGGCCCCTCTTTCCTCCAGTGCCAAGAAGAAGGGCAGAGAAGCAGAGGTGTTGGCCAAGGACA
GTGCCAAGCTTGCCAAGGCCTTGCTGAGGGAGCGGAAACAGGCGCACCGCCGTGCCAGCAGGCTCAC
CAGCCAGACGCAAGCCACGCTCCAACAGGCGTCCCAGCAGGTGCTGGCGTCTGAAGCACGCAGACAG
GAGCTGGAGGAAGCTGAGCGGGTGGGTGCTGGGCTGAGCGAGATGGAGCAGCAGATCCGGGAATCGC
GTATCTCACTGGAGAAGGACATCGAGACCTTGTCAGAGCTGCTTGCCAGGCTGGGGTCGCTGGACAC
CCATCAAGCCCCAGCCCAGGCCCTGAACGAGACTCAGTGGGCACTAGAACGCCTGAGGCTGCAGCTG
GGCTCCCCGGGGTCCTTGCAGAGGAAACTCAGTCTGCTGGAGCAGGAATCCCAGCAGCAGGAGCTGC
AGATCCAGGGCTTCGAGAGTGACCTCGCCGAGATCCGCGCCGACAAACAGAACCTGGAGGCCATTCT
GCACAGCCTGCCCGAGAACTGTGCCAGCTGGCAGTGA<u>GGGCTGCCCAGATCCCCGGCACACACTCCC</u>
<u>CCACCTGCTGTTTACATGACCCAGGGGGTGCACACTACCCCACAGGTGTGCCCATACAGACATTCCC</u>
<u>CGGAGCCGGCTGCTGTGAACTCGACCCCGTGTGGATAGTCACACTCCCTGCCGATTCTGTCTGTGGC</u>
<u>TTCTTCCCTGCCAGCAGGACTGAGTGTGCGTACCCAGTTCACCTGGACATGAGTGCACACTCTCACC</u>
<u>CCTGCACATGCATAAACGGCACACCCCAGTGTCAATAACATACACACGTGAGGTGCATGTCTGTG</u>
<u>TGTATGACCCAAATAAAAAAAAAAA</u>

	ORF Start: ATG at 98		ORF Stop: TGA at 4859

	SEQ ID NO: 210	1587 aa	MW at 172049.3kD
NOV49a, CG57292-01 Protein Sequence	MAAAALLLGLALLAPRAAGAGMGACYDGAGRPQRCLPVFENAAFGRLAQASHTCGSPPEDFCPHVGA AGAGAHCQRCDAADPQRHHNASYLTDFHSQDESTWWQSPSMAFGVQYPTSVNITLRLGKAYEITYVR LKFHTSRPESFAIYKRSRADGPWEPYQFYSASCQKTYGRPEGQYLRPGEDERVAFCTSEFSDISPLS GGNVAFSTLEGRPSAYNFEESPGLQEWVTSTELLISLDRLNTFGDDIFKDPKVLQSYYYAVSDFSVG GRCKCNGHASECGPDVAGQLACRCQHNTTGTDCERCLPFFQDRPWARGTAEAAHECLPCNCSGRSEE CTFDRELFRSTGHGGRCHHCRDHTAGPHCERCQENFYHWDPRMPCQPCDCQSAGSLHLQCDDTGTCA CKPTVTGWKCDRCLPGFHSLSEGGCRPCTCNPAGSLDTCDPRSGRCPCKENVEGNLCDRCRPGTFNL QPHNPAGCSSCFCYGHSKVCASTAQFQVHHILSDFHQGAEGWWARSVGGSEHSPQWSPNGVLLSPED EEELTAPGKFLGDQRFSYGQPLILTFRVPPGDSPLPVQLRLEGTGLALSLRHSSLSGPQDARASQGG RAQVPLQETSEDVAPPLPPFHFQRLLANLTSLRLRVSPGPSPAGPVFLTEVRLTSARPGLSPPASWV EICSCPTGYTGQFCESCAPGYKREMPQGGPYASCVPCTCNQHGTCDPNTGICVCSHHTEGPSCERCL PGFYGNPFAGQADDCQPCPCPGQSACTTIPESGEVVCTHCPPGQRGRRCEVCDDGFFGDPLGLFGHP QPCHQCQCSGNVDPNAVGNCDPLSGHCLRCLHNTTGDHCEHCQEGFYGSALAPRPADKCMPCSCHPQ GSVSEQMPCDPVTGQCSCLPHVTARDCSRCYPGFFDLQPGRGCRSCKCHPLGSQEDQCHPKTGQCTC RPGVTGQACDRCQLGFFGSSIKGCRACRCSPLGAASAQCHYNGTCVCRPGFEGYKCDRCHYNFFLTA DGTHCQQCPSCYALVKEETAKLKARLTLTEGWLQGSDCGSPWGPLDILLGEAPRGDVYQGHHLLPGA REAFLEQMMGLEGAVKAAREQLQRLNKGARCAQAGSQKTCTQLADLEAVLESSEEEILHAAAILASL EIPQEGPSQPTKWSHLAIEARALARSHRDTATKIAATAWRALLASNTSYALLWNLLEGRVALETQRD LEDRYQEVQAAQKALRTAVAEVLPEAESVLATVQQVGADTAPYLALLASPGALPQKSRAEDLGLKAK ALEKTVASWQHMATEAARTLQTAAQATLRQTEPLTMARSRLTATFASQLHQGARAALTQASSSVQAA TVTVMGARTLLADLEGMKLQFPRPKDQAALQRKADSVSDRLLADTRKKTKQAERMLGNAAPLSSSAK KKGREAEVLAKDSAKLAKALLRERKQAHRRASRLTSQTQATLQQASQQVLASEARRQELEEAERVGA GLSEMEQQIRESRISLEKDIETLSELLARLGSLDTHQAPAQALNETQWALERLRLQLGSPGSLQRKL SLLEQESQQQELQIQGFESDLAEIRADKQNLEAILHSLPENCASWQ		

	SEQ ID NO: 211	5148 bp	
NOV49b, CG57292-02 DNA Sequence	CCCCGCAGGGGAAGGCGGGTCCTGGCGGCCAGCGCGCGGTCCGCGCCCACCCTAGCCGACGGGGCCG GCAGAGCGCGCGGCGTCGGTGCCCTTGACC**ATG**GCGGCGGCTGCGCTTCTGCTGGGGCTGGCGCTGC TGGCACCGCGGGCGGCCGGCGCGGGCATGGGCGCGTGCTATGACGGCGCAGGGCGCCCGCAGCGCTG CCTGCCGGTGTTCGAGAACGCGGCGTTTGGGCGGCTCGCCCAGGCCTCGCACACGTGCGGCAGCCCG CCCGAGGACTTCTGTCCCCACGTGGGCGCCGCGGGCGCGGGGGCTCATTGCCAGCGCTGCGACGCCG CCGACCCCCAGCGCCACCACAACGCCTCCTACCTCACCGACTTCCACAGCCAGGACGAGAGCACCTG GTGGCAGAGCCCGTCCATGGCCTTCGGCGTGCAGTACCCCACCTCGGTCAACATCACCCTCCGCCTA GGGAAGGCTTATGAGATCACGTATGTGAGGCTGAAGTTCCACACCAGTCGCCCTGAGAGCTTTGCCA TCTACAAGCGCAGCCGCGCCGACGGCCCATGGAGCCCTACCAGTTCTACAGCGCCTCCTGCCAGAA GACCTACGGCCGGCCCGAGGGCCAGTACCTGCGCCCCGGCGAGGACGAGCGCGTGGCCTTCTGCACC TCTGAGTTCAGCGACATCTCCCCGCTGAGTGGCGGCAACGTGGCCTTCTCCACCCTGGAGGGCCGGC CCAGCGCCTACAACTTCGAGGAGAGCCCTGGGCTGCAGGAGTGGGTCACCAGCACCGAACTCCTCAT CTCTCTAGACCGGCTCAACACGTTTGGGGACGACATCTTCAAGGACCCCAAGGTGCTCCAGTCCTAC TATTATGCCGTGTCCGACTTCTCTGTGGGCGGCAGGTGCAAGTGCAACGGGCATGCCAGCGAGTGCG GCCCCGACGTGGCAGGCCAGTTGGCCTGCCGGTGCCAGCACAACACCACCGGCACAGACTGTGAGCG CTGCCTGCCCTTCTTCCAGGACCGCCCGTGGGCCCGGGGCACCGCCGAGGCTGCCCACGAGTGTCTG CCCTGCAACTGCAGTGGCCGCTCCGAGGAATGCACGTTTGATCGGGAGCTCTTCCGCAGCACAGGCC ACGGCGGGCGCTGTCACCACTGCCGTGACCACACAGCTGGGCCACACTGTGAGCGCTGTCAGGAGAA TTTCTATCACTGGGACCCGCGGATGCCATGCCAGCCCTGTGACTGCCAGTCGGCAGGCTCCCTACAC CTCCAGTGCGATGACACAGGCACCTGCGCCTGCAAGCCCACAGTGACTGGCTGGAAGTGTGACCGCT GTCTGCCCGGGTTCCACTCGCTCAGTGAGGGAGGCTGCAGACCCTGCACTTGCAATCCCGCTGGCAG CCTGGACACCTGTGACCCCCGCAGTGGGCGCTGCCCCTGCAAAGAGAATGTGGAAGGCAACCTATGT GACAGATGTCGCCCGGGGACCTTTAACCTGCAGCCCCACAATCCAGCTGGCTGCAGCAGCTGTTTCT GCTATGGCCACTCCAAGGTGTGCGCGTCCACTGCCCAGTTCCAGGTGCATCACATCCTCAGCGATTT CCACCAGGGAGCCGAAGGCTGGTGGGCCAGAAGTGTGGGGGGCTCTGAGCACTCCCCACAATGGAGC CCAAATGGGGTCCTCCTGAGCCCAGAAGACGAGGAGGAGCTCACAGCACCAGGGAAGTTCCTGGGAG ACCAGCGGTTCAGCTATGGGCAGCCCCTCATACTGACCTTCCGGGTGCCCCCCGGGGACTCCCCACT CCCTGTACAGCTGAGGCTGGAAGGACAGGCTTGGCCCTGTCCCTGAGGCACTCTAGCCTGTCTGGC CCCCAGGATGCCAGGGCATCCCAGGGAGGTAGAGCTCAGGTTCCACTGCAGGAGACCTCCGAGGACG TGGCCCCTCCACTGCCCCCCTTCCACTTCCAGCGGCTCCTCGCCAACCTGACCAGCCTCCGCCTCCG CGTCAGTCCCGGCCCCAGCCCTGCCGGTCCAGTGTTCCTGACTGAGGTCCGGCTCACATCCGCCCGG CCAGGGCTTTCCCCGCCAGCCTCCTGGGTGGAGATTTGTTCATGTCCCACTGGCTACACGGGCCAGT TCTGTGAATCCTGTGCTCCGGGATACAAGAGGGAGATGCCACAGGGGGGTCCCTATGCCAGCTGTGT CCCCTGCACCTGTAACCAGCATGGCACCTGTGACCCCAACACAGGGATCTGTGTCTGCAGCCACCAT ACCGAGGGCCCATCCTGTGAACGCTGTTTGCCAGGTTTCTATGGCAACCCTTTCGCGGGCCAAGCCG ACGACTGCCAGCCCTGTCCCTGCCCTGGCCAGTCGGCCTGTACGACCATCCCAGAGAGCGGGGAGGT GGTGTGTACCCACTGCCCCCCGGGCCAGAGAGGGCGGCGCTGTGAGGTCTGTGATGATGGCTTTTTT GGGGACCCGCTGGGGCTCTTTGGGCACCCCCAGCCCTGCCACCAGTGCCAGTGTAGCGGGAACGTGG ACCCCAATGCCGTGGGCAACTGTGACCCCCTGTCTGGCCACTGCCTGCGCTGCCTGCACAACACCAC GGGTGACCACTGTGAGCACTGTCAGGAAGGCTTCTACGGGAGCGCCCTGGCCCCTCGACCCGCAGAC AAATGCATGCCTTGCAGCTGTCACCCACAGGGCTCGGTCAGTGAGCAGATGCCCTGCGACCCAGTGA CAGGCCAATGCTCCTGCCTGCCTCATGTGACTGCACGGGACTGCAGCCGCTGCTACCCTGGCTTCTT CGACCTCCAGCCTGGAGGGGCTGCCGGAGCTGCAAGTGTCACCCACTGGGCTCCCAGGAGGACCAG		

	TGCCATCCCAAGACTGGACAGTGCACCTGCCGCCCAGGTGTCACAGGCCAGGCCTGTGACAGGTGCC AGCTGGGTTTCTTCGGCTCCTCAATCAAGGGCTGCCGGGCCTGCAGGTGCTCCCCACTGGGCGCTGC CTCGGCCCAGTGCCACTATAACGGCACATGCGTGTGCAGGCCTGGCTTCGAGGGCTACAAATGTGAC CGCTGCCACTACAACTTCTTCCTCACGGCAGACGGCACACACTGCCAGCAATGTCCGTCCTGCTACG CCCTGGTGAAGGAGGAGACAGCCAAGCTGAAGGCCAGACTGACTTTGACGGAGGGGTGGCTCCAAGG GTCCGACTGTGGCAGTCCCTGGGGACCACTAGACATTCTGCTGGGAGAGGCCCCAAGGGGGGACGTC TACCAGGGCCATCACCTGCTTCCAGGGGCTCGGGAAGCCTTCCTGGAGCAGATGATGGGCCTCGAGG GTGCTGTCAAGGCCGCCCGGGAGCAGCTGCAGAGGCTGAACAAGGGTGCCCGCTGTGCCCAGGCCGG ATCCCAGAAGACCTGCACCCAGCTGGCAGACCTGGAGGCAGTGCTGGAGTCCTCGGAAGAGGAGATT CTGCATGCAGCTGCCATTCTCGCGTCTCTGGAGATTCCTCAGGAAGGTCCCAGTCAGCCGACCAAAT GGAGCCACCTGGCCATAGAGGCCCGTGCCCTCGCCAGGAGCCACAGAGACACCGCCACCAAGATCGC AGCCACTGCTTGGAGGGCCCTGCTCGCCTCCAACACCAGCTACGCGCTTCTCTGGAATCTGCTGGAG GGAAGGGTGGCCCTAGAGACCCAGCGGGACCTGGAGGACAGGTACCAGGAGGTCCAGGCGGCCCAGA AAGCACTGAGGACGGCTGTGGCAGAGGTGCTGCCTGAAGCGGAAAGCGTGTTGGCCACCGTGCAGCA AGTTGGCGCAGATACAGCCCCGTACCTGGCCTTGCTGGCTTCCCCGGGAGCTCTGCCTCAGAAGTCC CGGGCTGAAGACCTGGGCCTGAAGGCGAAGGCCCTGGAGAAGACAGTTGCATCATGGCAGCACATGG CCACTGAGGCTGCCCGAACCCTCCAGACTGCTGCCCAGGCGACGCTACGGCAAACAGAACCCCTCAC AAAGCTGCACCAGGAGGCCAGAGCCGCCCTGACCCAGGCTTCCTCATCTGTCCAGGCTGCGACAGTG ACTGTCATGGGAGCCAGGACTCTGCTGGCTGATCTGGAAGGAATGAAGCTGCAGTTTCCCCGGCCCA AGGACCAGGCGGCATTGCAGAGGAAGGCAGACTCCGTCAGTGACAGACTCCTTGCAGACACGAGAAA GAAGACCAAGCAGGCGGAGAGGATGCTGGGAAACGCGGCCCCTCTTTCCTCCAGTGCCAAGAAGAAG GGCAGAGAAGCAGAGGTGTTGGCCAAGGACAGTGCCAAGCTTGCCAAGGCCTTGCTGAGGGAGCGGA AACAGGCGCACCGCCGTGCCAGCAGGCTCACCAGCCAGACGCAAGCCACGCTCCAACAGGCGTCCCA GCAGGTGCTGGCGTCTGAAGCACGCAGACAGGAGCTGGAGGAAGCTGAGCGGGTGGGTGCTGGGCTG AGCGAGATGGAGCAGCAGATCCGGGAATCGCGTATCTCACTGGAGAAGGACATCGAGACCTTGTCAG AGCTGCTTGCCAGGCTGGGGTCGCTGGACACCCATCAAGCCCCAGCCCAGGCCCTGAACGAGACTCA GTGGGCACTAGAACGCCTGAGGCTGCAGCTGGGCTCCCCGGGGTCCTTGCAGAGGAAACTCAGTCTG CTGGAGCAGGAATCCCAGCAGCAGGAGCTGCAGATCCAGGGCTTCGAGAGTGACCTCGCCGAGATCC GCGCCGACAAACAGAACCTGGAGGCCATTCTGCACAGCCTGCCCGAGAACTGTGCCAGCTGGCAGTG AGGGCTGCCCAGATCCCCGGCACACACTCCCCCACCTGCTGTTTACATGACCCAGGGGGTGCACACT ACCCCACAGGTGTGCCCATACAGACATTCCCCGGAGCCGGCTGCTGTGAACTCGACCCCGTGTGGAT AGTCACACTCCCTGCCGATTCTGTCTGTGGCTTCTTCCCTGCCAGCAGGACTGAGTGTGCGTACCCA GTTCACCTGGACATGAGTGCACACTCTCACCCCTGCACATGCATAAACGGGCACACCCCAGTGTCAA TAACATACACACGTGAGGGTGCATGTCTGTGTGTATGACCCAAATAAAAAAAAAAA		
	ORF Start: ATG at 98		ORF Stop: TGA at 4823

	SEQ ID NO: 212	1575 aa	MW at 170827.9kD
NOV49b, CG57292-02 Protein Sequence	MAAAALLLGLALLAPRAAGAGMGACYDGAGRPQRCLPVFENAAFGRLAQASHTCGSPPEDFCPHVGA AGAGAHCQRCDAADPQRHHNASYLTDFHSQDESTWWQSPSMAFGVQYPTSVNITLRLGKAYEITYVR LKFHTSRPESFAIYKRSRADGPWEPYQFYSASCQKTYGRPEGQYLRPGEDERVAFCTSEFSDISPLS GGNVAFSTLEGRPSAYNFEESPGLQEWVTSTELLISLDRLNTFGDDIFKDPKVLQSYYYAVSDFSVG GRCKCNGHASECGPDVAGQLACRCQHNTTGTDCERCLPFFQDRPWARGTAEAAHECLPCNCSGRSEE CTFDRELFRSTGHGGRCHHCRDHTAGPHCERCQENFYHWDPRMPCQPCDCQSAGSLHLQCDDTGTCA CKPTVTGWKCDRCLPGFHSLSEGGCRPCTCNPAGSLDTCDPRSGRCPCKENVEGNLCDRCRPGTFNL QPHNPAGCSSCFCYGHSKVCASTAQFQVHHILSDFHQGAEGWWARSVGGSEHSPQWSPNGVLLSPED EEELTAPGKFLGDQRFSYGQPLILTFRVPPGDSPLPVQLRLEGTGLALSLRHSSLSGPQDARASQGG RAQVPLQETSEDVAPPLPPFHFQRLLANLTSLRLRVSPGPSPAGPVFLTEVRLTSARPGLSPPASWV EICSCPTGYTGQFCESCAPGYKREMPQGGPYASCVPCTCNQHGTCDPNTGICVCSHHTEGPSCERCL PGFYGNPFAGQADDCQPCPCPGQSACTTIPESGEVVCTHCPPGQRGRRCEVCDDGFFGDPLGLFGHP QPCHQCQCSGNVDPNAVGNCDPLSGHCLRCLHNTTGDHCEHCQEGFYGSALAPRPADKCMPCSCHPQ GSVSEQMPCDPVTGQCSCLPHVTARDCSRCYPGFFDLQPGRGCRSCKCHPLGSQEDQCHPKTGQCTC RPGVTGQACDRCQLGFFGSSIKGCRACRCSPLGAASAQCHYNGTCVCRPGFEGYKCDRCHYNFFLTA DGTHCQQCPSCYALVKEETAKLKARLTLTEGWLQGSDCGSPWGPLDILLGEAPRGDVYQGHHLLPGA REAFLEQMMGLEGAVKAAREQLQRLNKGARCAQAGSQKTCTQLADLEAVLESSEEEILHAAAILASL EIPQEGPSQPTKWSHLAIEARALARSHRDTATKIAATAWRALLASNTSYALLWNLLEGRVALETQRD LEDRYQEVQAAQKALRTAVAEVLPEAESVLATVQQVGADTAPYLALLASPGALPQKSRAEDLGLKAK ALEKTVASWQHMATEAARTLQTAAQATLRQTEPLTKLHQEARAALTQASSSVQAATVTVMGARTLLA DLEGMKLQFPRPKDQAALQRKADSVSDRLLADTRKKTKQAERMLGNAAPLSSSAKKKGREAEVLAKD SAKLAKALLRERKQAHRRASRLTSQTQATLQQASQQVLASEARRQELEEAERVGAGLSEMEQQIRES RISLEKDIETLSELLARLGSLDTHQAPAQALNETQWALERLRLQLGSPGSLQRKLSLLEQESQQQEL QIQGFESDLAEIRADKQNLEAILHSLPENCASWQ		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 49B.

Table 49B. Comparison of NOV49a against NOV49b.

Protein Sequence	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV49b	25..1587 25..1575	1417/1563 (90%) 1418/1563 (90%)

Further analysis of the NOV49a protein yielded the following properties shown in Table 49C.

Table 49C. Protein Sequence Properties NOV49a	
PSort analysis:	0.5517 probability located in outside; 0.1900 probability located in lysosome (lumen); 0.1080 probability located in nucleus; 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 20 and 21

A search of the NOV49a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 49D.

Table 49D. Geneseq Results for NOV49a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM50361	Mouse laminin-15 gamma 3 chain - Mus musculus, 1587 aa. [WO200183516-A1, 08-NOV-2001]	1..1587 1..1587	1587/1587 (100%) 1587/1587 (100%)	0.0
AAB40917	Human ORFX ORF681 polypeptide sequence SEQ ID NO:1362 - Homo sapiens, 1587 aa. [WO200058473-A2, 05-OCT-2000]	1..1587 1..1587	1585/1587 (99%) 1586/1587 (99%)	0.0
AAY15458	Human laminin gamma 3 subunit - Homo sapiens, 1524 aa. [WO9919348-A1, 22-APR-1999]	67..1587 1..1524	1493/1524 (97%) 1496/1524 (97%)	0.0
AAB19803	Human laminin 2 gamma-1 chain with C-terminal FLAG epitope - Homo sapiens, 1617 aa. [WO200066730-A2, 09-NOV-2000]	10..1583 21..1600	698/1599 (43%) 964/1599 (59%)	0.0

AAB19801	Human laminin 2 gamma-1 chain - Homo sapiens, 1609 aa. [WO200066730-A2, 09-NOV-2000]	10..1583 21..1600	698/1599 (43%) 964/1599 (59%)	0.0

In a BLAST search of public sequence datbases, the NOV49a protein was found to have homology to the proteins shown in the BLASTP data in Table 49E.

Table 49E. Public BLASTP Results for NOV49a

Protein Accession Number	Protein/Organism/Length	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y6N6	Laminin gamma-3 chain precursor (Laminin 12 gamma 3) - Homo sapiens (Human), 1587 aa.	1..1587 1..1587	1587/1587 (100%) 1587/1587 (100%)	0.0
Q9R0B6	Laminin gamma-3 chain precursor (Laminin 12 gamma 3) - Mus musculus (Mouse), 1581 aa.	17..1585 26..1581	1169/1572 (74%) 1296/1572 (82%)	0.0
CAC17325	Sequence 25 from Patent WO0066730 precursor - Homo sapiens (Human), 1617 aa.	10..1583 21..1600	698/1599 (43%) 964/1599 (59%)	0.0
CAC17323	Sequence 21 from Patent WO0066730 precursor - Homo sapiens (Human), 1609 aa.	10..1583 21..1600	698/1599 (43%) 964/1599 (59%)	0.0
P11047	Laminin gamma-1 chain precursor (Laminin B2 chain) - Homo sapiens (Human), 1609 aa.	10..1583 21..1600	697/1599 (43%) 963/1599 (59%)	0.0

PFam analysis predicts that the NOV49a protein contains the domains shown in the Table 49F.

Table 49F. Domain Analysis of NOV49a

Pfam Domain	NOV49a Match Region	Identities/ Similarities for the Matched Region	Expect Value
laminin_Nterm	35..269	108/264 (41%) 182/264 (69%)	2.3e-110

laminin_EGF	271..324	17/63 (27%) 43/63 (68%)	3.4e-10
laminin_EGF	327..380	18/61 (30%) 49/61 (80%)	2e-13
laminin_EGF	383..427	27/59 (46%) 35/59 (59%)	5e-11
laminin_EGF	430..477	28/61 (46%) 46/61 (75%)	1.8e-14
laminin_B	541..671	44/152 (29%) 86/152 (57%)	1.6e-09
laminin_EGF	707..752	22/60 (37%) 39/60 (65%)	2.3e-12
laminin_EGF	755..807	17/61 (28%) 32/61 (52%)	0.0069
laminin_EGF	810..863	18/61 (30%) 45/61 (74%)	2.2e-15
laminin_EGF	866..914	27/60 (45%) 43/60 (72%)	1.3e-16
laminin_EGF	917..962	24/59 (41%) 39/59 (66%)	2e-15
laminin_EGF	965..1013	18/59 (31%) 37/59 (63%)	2e-07

Example 50.

The NOV50 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 50A.

Table 50A. NOV50 Sequence Analysis			
	SEQ ID NO: 213	942 bp	
NOV50a, CG97715-01 DNA Sequence	CCCCGGCTGCTTCTGCTCTTTCTGGTTCCGCTGCTGTGGGCCCCGGCTGCGGTCCGGGCCGGCCCAG ATGAAGACCTTAGCCACCGGAACAAAGAACCGCCGGCGCCGGCCCAGCAGCTGCAGCCGCAGCCTGT GGCTGTGCAGGGCCCCGAGCCGGCCCGGGTCGAGAAAATATTTACACCAGCAGCTCCAGTTCATACC AATAAAGAAGATCCTGCTACCCAAACTAATTTGGATTTATCCATGCATTTGTCGCTGCCATATCAG TTATTATTGTATCTGAATTGGGTGATAAGACATTTTTTATAGCAGCCATCATGGCAATGCGCTATAA CCGCCTGACCGTGCTGGCTGGTGCAATGCTTGCCTTGGGACTAATGACATGCTTGTCAGTTTTGTTT GGCTATGCCACCACAGTCATCCCCAGGGTCTATACATACTATGTTTCAACTGTATTATTTGCCATTT TTGGCATTAGAATGCTTCGGGAAGGCTTAAAGATGAGCCCTGATGAGGGTCAAGAGGAACTGGAAGA AGTTCAAGCTGAATTAAAGAAGAAAGATGAAGAATTTCAACGAACCAAACTTTTAAATGGACCGGGA GATGTTGAAACGGGTACAAGCATAACAGTACCTCAGAAAAAGTGGTTGCATTTTATTTCACCCATTT TTGTTCAAGCTCTTACATTAACATTCTTAGCAGAATGGGGTGATCGCTCTCAACTAACTACAATTGT ATTGGCAGCTAGAGAGGACCCCTATGGTGTAGCCGTGGGTGGAACTGTGGGGCACTGCCTGTGCACG GGATTGGCAGTAATTGGAGGAAGAATGATAGCACAGAAAATCTCTGTCAGAACTGTGACAATCATAG GAGGCATCGTTTTTTTGGCGTTTGCATTTTCTGCACTATTTATAAGGCCTGATTCTGGTTTTTAACA AGCT		
	ORF Start: at 1		ORF Stop: TAA at 934

	SEQ ID NO: 214	311 aa	MW at 33848.2kD
NOV50a, CG97715-01 Protein Sequence	PRLLLLFLVPLLWAPAAVRAGPDEDLSHRNKEPPAPAQQLQPQPVAVQGPEPARVEKIFTPAAPVHT NKEDPATQTNLGFIHAFVAAISVIIVSELGDKTFFIAAIMAMRYNRLTVLAGAMLALGLMTCLSVLF GYATTVIPRVYTYYVSTVLFAIFGIRMLREGLKMSPDEGQEELEEVQAELKKKDEEFQRTKLLNGPG DVETGTSITVPQKKWLHFISPIFVQALTLTFLAEWGDRSQLTTIVLAAREDPYGVAVGGTVGHCLCT GLAVIGGRMIAQKISVRTVTIIGGIVFLAFAFSALFIRPDSGF		

Further analysis of the NOV50a protein yielded the following properties shown in Table 50B.

Table 50B. Protein Sequence Properties NOV50a	
PSort analysis:	0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.3700 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 21 and 22

A search of the NOV50a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 50C.

Table 50C. Geneseq Results for NOV50a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB90211	Human polypeptide SEQ ID NO 2587 - Homo sapiens, 324 aa. [WO200190304-A2, 29-NOV-2001]	1..311 14..324	310/311 (99%) 310/311 (99%)	e-176
AAB51239	Human hTMPT27 protein sequence SEQ ID NO:7 - Homo sapiens, 324 aa. [CN1268567-A, 04-OCT-2000]	1..311 14..324	310/311 (99%) 310/311 (99%)	e-176
AAB20092	Human hydrophobic domain-containing protein HP03373 - Homo sapiens, 324 aa. [WO200100824-A2, 04-JAN-2001]	1..311 14..324	310/311 (99%) 310/311 (99%)	e-176
AAB41971	Human ORFX ORF1735 polypeptide sequence SEQ ID NO:3470 - Homo sapiens, 322 aa. [WO200058473-A2, 05-OCT-2000]	1..311 12..322	310/311 (99%) 310/311 (99%)	e-176

ABB57033	Mouse ischaemic condition related protein sequence SEQ ID NO:39 - Mus musculus, 323 aa. [WO200188188-A2, 22-NOV-2001]	2..311 15..323	282/310 (90%) 288/310 (91%)	e-158

In a BLAST search of public sequence datbases, the NOV50a protein was found to have homology to the proteins shown in the BLASTP data in Table 50D.

Table 50D. Public BLASTP Results for NOV50a

Protein Accession Number	Protein/Organism/Length	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9HC07	Transmembrane protein PT27 - Homo sapiens (Human), 324 aa.	1..311 14..324	310/311 (99%) 310/311 (99%)	e-176
Q9NZ34	Uncharacterized hypothalamus protein HTMP - Homo sapiens (Human), 324 aa.	1..311 14..324	310/311 (99%) 310/311 (99%)	e-176
Q9R292	TPARDL - Mus musculus (Mouse), 323 aa.	2..311 15..323	287/310 (92%) 293/310 (93%)	e-161
P52875	Transmembrane protein PFT27 (TPA regulated locus protein) - Mus musculus (Mouse), 323 aa.	2..311 15..323	282/310 (90%) 288/310 (91%)	e-158
AAM21311	Transmembrane protein HTP-1 - Brachydanio rerio (Zebrafish) (Danio rerio), 305 aa.	41..311 34..305	216/272 (79%) 236/272 (86%)	e-118

PFam analysis predicts that the NOV50a protein contains the domains shown in the Table 50E.

Table 50E. Domain Analysis of NOV50a

Pfam Domain	NOV50a Match Region	Identities/ Similarities for the Matched Region	Expect Value
UPF0016	84..158	36/76 (47%) 74/76 (97%)	9.9e-39

UPF0016	224..299	42/76 (55%) 76/76 (100%)	8.1e-44

Example B: Sequencing Methodology and Identification of NOVX Clones

1. GeneCalling™ Technology: This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

2. SeqCalling™ Technology: cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using

CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

3. **PathCalling™ Technology:** The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corportion proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for

corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

4. **RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5. **Exon Linking:** The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain -

hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6. **Physical Clone:** Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

Example C: Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell

lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoinflammatory diseases), Panel CNSD.01 (containing samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β-actin and GAPDH). Normalized RNA (5 ul) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10 μg of total RNA were performed in a volume of 20 μl and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50 μg of total RNA in a final volume of 100 μl. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (Tm) range = 58°-60°C, primer optimal Tm = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe Tm must be 10°C greater than primer Tm, amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes

were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and

were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,
* = established from metastasis,
met = metastasis,
s cell var = small cell variant,
non-s = non-sm = non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,
glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

General_screening_panel_v1.4, v1.5 and v1.6

The plates for Panels 1.4, 1.5, and 1.6 include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panels 1.4, 1.5, and 1.6 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panels 1.4, 1.5, and 1.6 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panels 1.4, 1.5, and 1.6 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal

lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

Panels 2D, 2.2, 2.3 and 2.4

The plates for Panels 2D, 2.2, 2.3 and 2.4 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI) or from Ardais or Clinomics). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI/ CHTN/Ardais/Clinomics). Unmatched RNA samples from tissues without malignancy (normal tissues) were also obtained from Ardais or Clinomics. This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

HASS Panel v 1.0

The HASS panel v 1.0 plates are comprised of 93 cDNA samples and two controls. Specifically, 81 of these samples are derived from cultured human cancer cell lines that had been subjected to serum starvation, acidosis and anoxia for different time periods as well as controls for these treatments, 3 samples of human primary cells, 9 samples of malignant brain cancer (4 medulloblastomas and 5 glioblastomas) and 2 controls. The human cancer cell lines are obtained from ATCC (American Type Culture Collection) and fall into the following tissue groups: breast cancer, prostate cancer, bladder carcinomas, pancreatic

cancers and CNS cancer cell lines. These cancer cells are all cultured under standard recommended conditions. The treatments used (serum starvation, acidosis and anoxia) have been previously published in the scientific literature. The primary human cells were obtained from Clonetics (Walkersville, MD) and were grown in the media and conditions recommended by Clonetics. The malignant brain cancer samples are obtained as part of a collaboration (Henry Ford Cancer Center) and are evaluated by a pathologist prior to CuraGen receiving the samples . RNA was prepared from these samples using the standard procedures. The genomic and chemistry control wells have been described previously.

ARDAIS Panel v 1.0

The plates for ARDAIS panel v 1.0 generally include 2 control wells and 22 test samples composed of RNA isolated from human tissue procured by surgeons working in close cooperation with Ardais Corporation. The tissues are derived from human lung malignancies (lung adenocarcinoma or lung squamous cell carcinoma) and in cases where indicated many malignant samples have "matched margins" obtained from noncancerous lung tissue just adjacent to the tumor. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue) in the results below. The tumor tissue and the "matched margins" are evaluated by independent pathologists (the surgical pathologists and again by a pathologist at Ardais). Unmatched malignant and non-malignant RNA samples from lungs were also obtained from Ardais. Additional information from Ardais provides a gross histopathological assessment of tumor differentiation grade and stage. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical state of the patient.

Panel 3D, 3.1 and 3.2

The plates of Panel 3D, 3.1, and 3.2 are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard

recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D, 3.1, 3.2, 1, 1.1., 1.2, 1.3D, 1.4, 1.5, and 1.6 are of the most common cell lines used in the scientific literature.

Panels 4D, 4R, and 4.1D

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10^{-5}M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2μg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10^{-5}M (Gibco), and 10mM Hepes (Gibco)

with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately $2x10^6$cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol ($5.5x10^{-5}$M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10μg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and plated at 10^6cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5μg/ml anti-CD28 (Pharmingen) and 3ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated

CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10^{-5}M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10^{-5}M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resupended at 10^{6}cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10^{-5}M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5μg/ml or anti-CD40 (Pharmingen) at approximately 10μg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10μg/ml anti-CD28 (Pharmingen) and 2μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^{5}-10^{6}cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10^{-5}M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1μg/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1μg/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10^{-5}M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1μg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and

Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at $5x10^5$cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to $5x10^5$cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1μg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300μl of RNAse-free water and 35μl buffer (Promega) 5μl DTT, 7μl RNAsin and 8μl DNAse were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNAse free water. RNA was stored at -80°C.

AI_comprehensive panel_v1.0

The plates for AI_comprehensive panel_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues

obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebvid and two were on phenobarbital.

Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI_comprehensive panel_v1.0 panel, the following abbreviations are used:

AI = Autoimmunity

Syn = Synovial

Normal = No apparent disease

Rep22 /Rep20 = individual patients

RA = Rheumatoid arthritis

Backus = From Backus Hospital

OA = Osteoarthritis

(SS) (BA) (MF) = Individual patients

Adj = Adjacent tissue

Match control = adjacent tissues

-M = Male

-F = Female

COPD = Chronic obstructive pulmonary disease

Panels 5D and 5I

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

Patient 2: Diabetic Hispanic, overweight, not on insulin

Patient 7-9: Nondiabetic Caucasian and obese (BMI>30)

Patient 10: Diabetic Hispanic, overweight, on insulin

Patient 11: Nondiabetic African American and overweight

Patient 12: Diabetic Hispanic on insulin

Adiocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose

Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated

Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

GO Adipose = Greater Omentum Adipose

SK = Skeletal Muscle

UT = Uterus

PL = Placenta

AD = Adipose Differentiated

AM = Adipose Midway Differentiated

U = Undifferentiated Stem Cells

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by

neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

In the labels employed to identify tissues in the CNS_Neurodegeneration_V1.0 panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

Control (Path) = Control brains; pateint not demented but showing sever AD-like pathology

SupTemporal Ctx = Superior Temporal Cortex

Inf Temporal Ctx = Inferior Temporal Cortex

A. CG105472-01: KIAA0575/Greb1

Expression of gene CG105472-01 was assessed using the primer-probe sets Ag3041, Ag3042, Ag4301 and Ag4300, described in Tables AA, AB, AC and AD. Results of the RTQ-PCR runs are shown in Tables AE, AF, AG, AH, AI, AJ and AK.

Table AA. Probe Name Ag3041

Primers		Length	Start Position	SEQ ID No
Forward	5'-gtattacctggtccgtaatgca-3'	22	870	215
Probe	TET-5'-caagggactctaaccaaaggaccttt -3'-TAMRA	26	892	216
Reverse	5'-ggcttctaaactctgagccttt-3'	22	928	217

Table AB. Probe Name Ag3042

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-gtattacctggtccgtaatgca-3'	22	870	218
Probe	TET-5'-caagggactctaaccaaaggaccttt -3'-TAMRA	26	892	219
Reverse	5'-ggcttctaaactctgagccttt-3'	22	928	220

Table AC. Probe Name Ag4301

Primers	Sequencs	Length	Start Position	SEQ ID No
Forward	5'-ctgtggaaagaaaggcttctg-3'	21	777	221
Probe	TET-5'-tcacggaattctccaatcatataaatc tg-3'-TAMRA	29	803	222
Reverse	5'-cttgggttgagtggtcagttt-3'	21	832	223

Table AD. Probe Name Ag4300

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-gccaagtaggttcccctgta-3'	20	6545	224
Probe	TET-5'-cctcctacaaagcaatattccaaagg a-3'-TAMRA	27	6566	225
Reverse	5'-ttcttgtctccagcctttacag-3'	22	6602	226

Table AE. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag4300, Run 224063041	Rel. Exp.(%) Ag4301, Run 224064603	issue Name	Rel. Exp.(%) Ag4300, Run 224063041	Rel. Exp.(%) Ag4301, Run 224064603
AD 1 Hippo	32.8	15.4	Control (Path) 3 Temporal Ctx	30.1	10.8
AD 2 Hippo	59.0	28.9	Control (Path) 4 Temporal Ctx	48.0	42.0

AD 3 Hippo	20.6	10.0	AD 1 Occipital Ctx	31.6	24.3
AD 4 Hippo	22.4	9.2	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 Hippo	**100.0**	**100.0**	AD 3 Occipital Ctx	21.6	6.6
AD 6 Hippo	79.6	35.1	AD 4 Occipital Ctx	40.3	26.1
Control 2 Hippo	43.8	21.3	AD 5 Occipital Ctx	72.7	25.9
Control 4 Hippo	25.2	12.3	AD 6 Occipital Ctx	40.6	47.0
Control (Path) 3 Hippo	18.8	7.7	Control 1 Occipital Ctx	5.6	5.5
AD 1 Temporal Ctx	52.9	17.1	Control 2 Occipital Ctx	73.2	62.9
AD 2 Temporal Ctx	90.1	34.6	Control 3 Occipital Ctx	34.6	2.0
AD 3 Temporal Ctx	20.6	8.9	Control 4 Occipital Ctx	22.1	8.1
AD 4 Temporal Ctx	45.7	31.0	Control (Path) 1 Occipital Ctx	87.7	65.1
AD 5 Inf Temporal Ctx	89.5	71.7	Control (Path) 2 Occipital Ctx	28.5	27.0
AD 5 Sup Temporal Ctx	58.6	31.9	Control (Path) 3 Occipital Ctx	8.2	3.4
AD 6 Inf Temporal Ctx	59.0	29.7	Control (Path) 4 Occipital Ctx	31.6	33.0
AD 6 Sup Temporal Ctx	65.5	43.8	Control 1 Parietal Ctx	27.4	12.1
Control 1 Temporal Ctx	18.9	8.5	Control 2 Parietal Ctx	66.0	31.2
Control 2 Temporal Ctx	43.5	30.4	Control 3 Parietal Ctx	31.6	28.5
Control 3 Temporal Ctx	37.6	24.1	Control (Path) 1 Parietal Ctx	81.2	62.0
Control 3 Temporal Ctx	22.4	13.0	Control (Path) 2 Parietal Ctx	45.4	35.1
Control (Path) 1 Temporal Ctx	67.4	48.6	Control (Path) 3 Parietal Ctx	13.4	7.9
Control (Path) 2 Temporal Ctx	48.0	38.4	Control (Path) 4 Parietal Ctx	58.6	57.8

Table AF. General_screening_panel_v1.4

Tissue Name	**Rel. Exp.(%) Ag4300, Run 221998703**	**issue Name**	**Rel. Exp.(%) Ag4300, Run 221998703**
Adipose	0.8	Renal ca. TK-10	9.3
Melanoma* Hs688(A).T	0.1	Bladder	1.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.3

Melanoma* M14	3.8	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	3.9	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	45.7	Colon ca. SW480	0.5
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.4
Testis Pool	7.4	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	2.2
Prostate Pool	14.9	Colon ca. CaCo-2	1.4
Placenta	0.8	Colon cancer tissue	0.3
Uterus Pool	8.6	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	2.5	Colon ca. Colo-205	0.3
Ovarian ca. SK-OV-3	1.9	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	36.3
Ovarian ca. OVCAR-5	16.7	Small Intestine Pool	11.4
Ovarian ca. IGROV-1	1.5	Stomach Pool	12.3
Ovarian ca. OVCAR-8	1.6	Bone Marrow Pool	22.8
Ovary	24.8	Fetal Heart	1.6
Breast ca. MCF-7	**100.0**	Heart Pool	6.7
Breast ca. MDA-MB-231	0.2	Lymph Node Pool	48.3
Breast ca. BT 549	0.2	Fetal Skeletal Muscle	3.3
Breast ca. T47D	44.8	Skeletal Muscle Pool	1.3
Breast ca. MDA-N	11.9	Spleen Pool	0.5
Breast Pool	39.8	Thymus Pool	14.6
Trachea	0.9	CNS cancer (glio/astro) U87-MG	5.8
Lung	39.2	CNS cancer (glio/astro) U-118-MG	0.2
Fetal Lung	1.4	CNS cancer (neuro;met) SK-N-AS	3.1
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.5
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	1.6
Lung ca. NCI-H146	1.0	CNS cancer (glio) SNB-19	1.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	2.2
Lung ca. A549	1.3	Brain (Amygdala) Pool	2.4
Lung ca. NCI-H526	0.6	Brain (cerebellum)	2.9
Lung ca. NCI-H23	0.6	Brain (fetal)	2.4
Lung ca. NCI-H460	1.6	Brain (Hippocampus) Pool	3.4
Lung ca. HOP-62	1.6	Cerebral Cortex Pool	5.8
Lung ca. NCI-H522	3.3	Brain (Substantia nigra) Pool	2.8
Liver	0.1	Brain (Thalamus) Pool	5.6
Fetal Liver	4.6	Brain (whole)	3.7
Liver ca. HepG2	20.9	Spinal Cord Pool	4.3
Kidney Pool	21.5	Adrenal Gland	3.4
Fetal Kidney	3.6	Pituitary gland Pool	1.1
Renal ca. 786-0	0.0	Salivary Gland	0.1
Renal ca. A498	3.0	Thyroid (female)	0.0
Renal ca. ACHN	1.3	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	1.3	Pancreas Pool	25.2

Table AG. Panel 1.3D

Tissue Name	Rel. Exp.(%) g3041, Run 167964342	Rel. Exp.(%) Ag3042, Run 167964481	Tissue Name	Rel. Exp.(%) Ag3041, Run 167964342	Rel. Exp.(%) Ag3042, Run 167964481
Liver adenocarcinoma	0.1	0.1	Kidney (fetal)	2.1	1.6
Pancreas	0.1	0.3	Renal ca. 786-0	0.0	0.0
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. A498	0.0	0.1
Adrenal gland	0.6	0.5	Renal ca. RXF 393	0.0	0.0
Thyroid	0.0	0.0	Renal ca. ACHN	0.0	0.0
Salivary gland	0.0	0.0	Renal ca. UO-31	0.0	0.0
Pituitary gland	0.6	1.0	Renal ca. TK-10	0.0	0.0
Brain (fetal)	0.5	0.3	Liver	0.1	0.0
Brain (whole)	1.1	0.4	Liver (fetal)	0.3	0.5
Brain (amygdala)	0.6	0.6	Liver ca. (hepatoblast) HepG2	32.3	27.5
Brain (cerebellum)	1.4	1.0	Lung	0.1	0.0
Brain (hippocampus)	0.6	1.0	Lung (fetal)	0.1	0.2
Brain (substantia nigra)	0.7	0.6	Lung ca. (small cell) LX-1	0.0	0.0
Brain (thalamus)	0.2	0.1	Lung ca. (small cell) NCI-H69	0.5	0.3
Cerebral Cortex	4.4	4.4	Lung ca. (s.cell var.) SHP-77	0.5	0.3
Spinal cord	0.5	0.9	Lung ca. (large cell)NCI-H460	0.1	0.1
glio/astro U87-MG	0.0	0.0	Lung ca. (non-sm. cell) A549	0.3	0.1
glio/astro U-118-MG	0.0	0.0	Lung ca. (non-s.cell) NCI-H23	0.3	0.5
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cell) HOP-62	0.1	0.1
neuro*; met SK-N-AS	1.3	0.9	Lung ca. (non-s.cl) NCI-H522	1.8	1.6
astrocytoma SF-539	0.1	0.0	Lung ca. (squam.) SW 900	0.5	0.2
astrocytoma SNB-75	0.2	0.2	Lung ca. (squam.) NCI-H596	3.2	2.7
glioma SNB-19	0.0	0.0	Mammary gland	0.7	0.5
glioma U251	0.0	0.0	Breast ca.* (pl.ef) MCF-7	**100.0**	**100.0**

glioma SF-295	0.0	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
Heart (fetal)	2.2	1.0	Breast ca.* (pl.ef) T47D	91.4	90.8
Heart	0.3	0.3	Breast ca. BT-549	0.0	0.0
Skeletal muscle (fetal)	2.5	2.5	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	0.8	0.5	Ovary	46.0	46.7
Bone marrow	0.0	0.0	Ovarian ca. OVCAR-3	0.9	0.6
Thymus	0.0	0.0	Ovarian ca. OVCAR-4	0.1	0.1
Spleen	0.0	0.0	Ovarian ca. OVCAR-5	0.1	0.0
Lymph node	0.0	0.0	Ovarian ca. OVCAR-8	0.1	0.1
Colorectal	0.1	0.1	Ovarian ca. IGROV-1	0.9	0.6
Stomach	0.1	0.1	Ovarian ca.* (ascites) SK-OV-3	2.2	1.3
Small intestine	0.2	0.1	Uterus	7.1	6.2
Colon ca. SW480	0.1	0.1	Placenta	0.1	0.1
Colon ca.* SW620(SW480 met)	0.7	0.6	Prostate	7.6	7.7
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met)PC-3	0.0	0.0
Colon ca. HCT-116	0.5	0.5	Testis	1.0	0.8
Colon ca. CaCo-2	0.9	0.7	Melanoma Hs688(A).T	0.0	0.0
Colon ca. tissue(ODO3866)	0.1	0.1	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC-2998	0.6	0.3	Melanoma UACC-62	0.0	0.1
Gastric ca.* (liver met) NCI-N87	0.0	0.0	Melanoma M14	0.0	0.0
Bladder	0.5	0.4	Melanoma LOX IMVI	0.0	0.0
Trachea	0.0	0.0	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney	0.7	0.7	Adipose	0.4	0.3

Table AH. Panel 2.2

Tissue Name	**Rel. Ex.(%) Ag3041, Run 174441332**	**Tissue Name**	**Rel. Exp.(%) Ag3041, Run 174441332**

Normal Colon	6.3	Kidney Margin (OD04348)	5.2
Colon cancer (OD06064)	0.4	Kidney malignant cancer (OD06204B)	0.5
Colon Margin (OD06064)	0.3	Kidney normal adjacent tissue (OD06204E)	0.6
Colon cancer (OD06159)	0.0	Kidney Cancer (OD04450-01)	2.5
Colon Margin (OD06159)	0.3	Kidney Margin (OD04450-03)	1.0
Colon cancer (OD06297-04)	0.0	Kidney Cancer 8120613	0.0
Colon Margin (OD06297-05)	0.0	Kidney Margin 8120614	5.4
CC Gr.2 ascend colon (ODO3921)	0.0	Kidney Cancer 9010320	0.4
CC Margin (ODO3921)	0.4	Kidney Margin 9010321	1.0
Colon cancer metastasis (OD06104)	0.0	Kidney Cancer 8120607	0.0
Lung Margin (OD06104)	0.3	Kidney Margin 8120608	1.2
Colon mets to lung (OD04451-01)	0.0	Normal Uterus	54.0
Lung Margin (OD04451-02)	0.0	Uterine Cancer 064011	7.1
Normal Prostate	22.4	Normal Thyroid	0.0
Prostate Cancer (OD04410)	16.6	Thyroid Cancer 064010	0.0
Prostate Margin (OD04410)	24.1	Thyroid Cancer A302152	0.0
Normal Ovary	**100.0**	Thyroid Margin A302153	0.0
Ovarian cancer (OD06283-03)	1.3	Normal Breast	10.3
Ovarian Margin (OD06283-07)	0.0	Breast Cancer (OD04566)	1.2
Ovarian Cancer 064008	1.4	Breast Cancer 1024	17.1
Ovarian cancer (OD06145)	0.7	Breast Cancer (OD04590-01)	62.0
Ovarian Margin (OD06145)	17.9	Breast Cancer Mets (OD04590-03)	40.6
Ovarian cancer (OD06455-03)	3.8	Breast Cancer Metastasis (OD04655-05)	53.2
Ovarian Margin (OD06455-07)	10.4	Breast Cancer 064006	7.9
Normal Lung	0.1	Breast Cancer 9100266	87.7
Invasive poor diff. lung adeno (ODO4945-01	0.6	Breast Margin 9100265	43.2
Lung Margin (ODO4945-03)	0.3	Breast Cancer A209073	2.1
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	17.0
Lung Margin (OD03126)	0.0	Breast cancer (OD06083)	52.1
Lung Cancer (OD05014A)	0.0	Breast cancer node metastasis (OD06083)	40.9
Lung Margin (OD05014B)	0.3	Normal Liver	1.1
Lung cancer (OD06081)	0.4	Liver Cancer 1026	0.5
Lung Margin (OD06081)	0.1	Liver Cancer 1025	3.0
Lung Cancer (OD04237-01)	0.0	Liver Cancer 6004-T	0.0
Lung Margin (OD04237-02)	0.2	Liver Tissue 6004-N	0.0
Ocular Melanoma Metastasis	2.0	Liver Cancer 6005-T	1.6
Ocular Melanoma Margin (Liver)	0.9	Liver Tissue 6005-N	1.3

Melanoma Metastasis	0.1	Liver Cancer 064003	4.0
Melanoma Margin (Lung)	0.2	Normal Bladder	0.6
Normal Kidney	1.5	Bladder Cancer 1023	2.7
Kidney Ca, Nuclear grade 2 (OD04338)	1.4	Bladder Cancer A302173	0.0
Kidney Margin (OD04338)	1.7	Normal Stomach	1.2
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04339)	1.7	Stomach Margin 9060396	0.5
Kidney Ca, Clear cell type (OD04340)	0.0	Gastric Cancer 9060395	0.5
Kidney Margin (OD04340)	1.1	Stomach Margin 9060394	1.5
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 064005	0.2

Table AI. Panel 3D

Tissue Name	Rel. Exp.(%) Ag3041, Run 182098857	Rel. Exp.(%) Ag4300, Run 182114559	Tissue Name	Rel. Exp.(%) Ag3041, Run 182098857	Rel. Exp.(%) Ag4300, Run 182114559
Daoy- Medulloblastoma	11.9	0.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	2.7	0.0
TE671- Medulloblastoma	0.8	0.0	ES-2- Ovarian clear cell carcinoma	0.0	0.0
D283 Med- Medulloblastoma	9.4	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0	0.0
PFSK-1- Primitive Neuroectodermal	27.4	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0	0.0
XF-498- CNS	0.0	10.3	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	21.0	0.0
SNB-78- Glioma	0.0	0.0	Raji- Burkitt's lymphoma	0.9	0.0
SF-268- Glioblastoma	1.0	0.0	Daudi- Burkitt's lymphoma	2.0	0.0
T98G- Glioblastoma	0.0	0.0	U266- B-cell plasmacytoma	75.8	2.5
SK-N-SH- Neuroblastoma (metastasis)	**100.0**	7.9	CA46- Burkitt's lymphoma	5.6	0.0
SF-295- Glioblastoma	0.0	0.0	RL- non-Hodgkin's B-cell lymphoma	2.2	0.0

Cerebellum	22.4	0.0	JM1- pre-B-cell lymphoma	0.0	0.0
Cerebellum	19.8	0.0	Jurkat- T cell leukemia	1.0	0.0
NCI-H292- Mucoepidermoid lung carcinoma	0.0	0.0	TF-1- Erythroleukemia	0.0	0.0
DMS-114- Small cell lung cancer	1.8	0.0	HUT 78- T-cell lymphoma	40.6	0.0
DMS-79- Small cell lung cancer	32.1	**100.0**	U937- Histiocytic lymphoma	14.7	0.0
NCI-H146- Small cell lung cancer	24.1	0.0	KU-812- Myelogenous leukemia	0.9	0.0
NCI-H526- Small cell lung cancer	57.4	0.0	769-P- Clear cell renal carcinoma	2.9	0.0
NCI-N417- Small cell lung cancer	0.9	0.0	Caki-2- Clear cell renal carcinoma	0.0	0.0
NCI-H82- Small cell lung cancer	69.7	5.5	SW 839- Clear cell renal carcinoma	0.0	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	1.7	0.0	G401- Wilms' tumor	36.1	0.0
NCI-H1155- Large cell lung cancer	8.5	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	0.0	0.0
NCI-H1299- Large cell lung cancer	4.5	0.0	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	2.1	0.0
NCI-H727- Lung carcinoid	1.9	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	6.8	0.0
NCI-UMC-11- Lung carcinoid	2.2	0.0	BxPC-3- Pancreatic adenocarcinoma	0.0	0.0
LX-1- Small cell lung cancer	0.0	0.0	HPAC- Pancreatic adenocarcinoma	0.0	0.0
Colo-205- Colon cancer	9.2	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0	0.0
KM12- Colon cancer	2.0	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	3.3	0.0
KM20L2- Colon cancer	0.0	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0	0.0
NCI-H716- Colon cancer	0.9	0.0	T24- Bladder carcinma (transitional cell)	0.0	0.0
SW-48- Colon adenocarcinoma	0.0	0.0	5637- Bladder carcinoma	0.0	0.0
SW1116- Colon adenocarcinoma	0.0	0.0	HT-1197- Bladder carcinoma	0.0	0.0

LS 174T- Colon adenocarcinoma	0.0	0.0	UM-UC-3- Bladder carcinma (transitional cell)	0.0	0.0
SW-948- Colon adenocarcinoma	1.1	0.0	A204- Rhabdomyosarcoma	6.7	0.0
SW-480- Colon adenocarcinoma	0.0	0.0	HT-1080- Fibrosarcoma	0.0	5.4
NCI-SNU-5- Gastric carcinoma	3.9	0.0	MG-63- Osteosarcoma	7.7	0.0
KATO III- Gastric carcinoma	1.0	0.0	SK-LMS-1- Leiomyosarcoma (vulva)	0.0	0.0
NCI-SNU-16- Gastric carcinoma	0.0	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0	0.0
NCI-SNU-1- Gastric carcinoma	0.0	0.0	A431- Epidermoid carcinoma	0.0	0.0
RF-1- Gastric adenocarcinoma	1.9	0.0	WM266-4- Melanoma	0.0	94.6
RF-48- Gastric adenocarcinoma	0.0	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0	0.0
MKN-45- Gastric carcinoma	0.0	0.0	MDA-MB-468- Breast adenocarcinoma	0.0	0.0
NCI-N87- Gastric carcinoma	0.0	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0	0.0
OVCAR-5- Ovarian carcinoma	1.8	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0	0.0
RL95-2- Uterine carcinoma	0.0	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0	0.0
HelaS3- Cervical adenocarcinoma	8.0	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0	0.0

Table AJ. Panel 4.1D

Tissue Name	Rel. Exp.() Ag4301, Run 181981970	Tissue Name	Rel. Exp.(%) Ag4301, Run 181981970
Secondary Th1 act	0.2	HUVEC IL-1beta	0.0
Secondary Th2 act	1.0	HUVEC IFN gamma	1.1
Secondary Tr1 act	1.3	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.5	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.6	HUVEC IL-11	0.0
Secondary Tr1 rest	0.7	Lung Microvascular EC none	0.0

Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.9	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.5	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	1.8	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.8	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	4.6	Astrocytes rest	1.8
Secondary CD8 lymphocyte rest	0.4	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	2.1	KU-812 (Basophil) rest	0.4
CD4 lymphocyte none	0.6	KU-812 (Basophil) PMA/ionomycin	0.6
2ry Th1/Th2/Tr1_anti-CD95 CH11	1.9	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.6
LAK cells IL-2	3.1	Liver cirrhosis	1.5
LAK cells IL-2+IL-12	2.1	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.9	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.6
NK Cells IL-2 rest	2.0	NCI-H292 IFN gamma	0.3
Two Way MLR 3 day	0.6	HPAEC none	0.0
Two Way MLR 5 day	0.8	HPAEC TNF alpha + IL-1 beta	0.4
Two Way MLR 7 day	1.3	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	1.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.7	Lung fibroblast IL-9	0.0
Ramos (B cell) none	1.3	Lung fibroblast IL-13	1.3
Ramos (B cell) ionomycin	3.8	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.6	Dermal fibroblast CCD1070 rest	0.6
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	1.4
EOL-1 dbcAMP	**100.0**	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.3	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.3	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0

Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.7
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.4	Colon	0.9
Macrophages rest	0.0	Lung	5.1
Macrophages LPS	0.0	Thymus	5.8
HUVEC none	0.0	Kidney	54.3
HUVEC starved	0.0		

Table AK. general oncology screening panel_v_2.4

Tissue Name	Rel. Exp.(%) Ag3042, Run 268695244	Rel. Exp.(%) Ag4300, Run 260280468	Rel. Exp.(%) Ag4301, Run 268665966	issue Name	Rel. Exp.(%) Ag3042, Run 268695244	Rel. Exp.(%) Ag4300, Run 260280468	Rel. Exp.(%) Ag4301, Run 268665966
Colon cancer 1	0.1	0.0	0.1	Bladder NAT 2	0.0	0.0	0.0
Colon NAT 1	0.1	0.0	0.0	Bladder NAT 3	0.0	0.0	0.0
Colon cancer 2	0.2	0.0	0.1	Bladder NAT 4	0.1	0.0	0.0
Colon NAT 2	0.0	0.0	0.0	Prostate adenocarcinoma 1	1.7	5.6	1.1
Colon cancer 3	0.1	0.0	0.1	Prostate adenocarcinoma 2	2.6	0.8	1.1
Colon NAT 3	0.1	0.0	0.1	Prostate adenocarcinoma 3	20.0	31.9	13.4
Colon malignant cancer 4	0.1	0.0	0.1	Prostate adenocarcinoma 4	0.7	0.0	0.7
Colon NAT 4	0.0	0.0	0.0	Prostate NAT 5	4.3	6.5	2.5
Lung cancer 1	0.0	0.0	0.0	Prostate adenocarcinoma 6	11.1	14.5	9.2
Lung NAT 1	0.0	0.0	0.0	Prostate adenocarcinoma 7	7.0	9.9	3.3
Lung cancer 2	6.9	14.7	5.8	Prostate adenocarcinoma 8	2.9	4.0	1.6
Lung NAT 2	0.0	0.0	0.0	Prostate adenocarcinoma 9	7.9	12.5	5.8

Squamous cell carcinoma 3	0.3	0.0	0.2	Prostate NAT 10	2.7	4.7	2.0
Lung NAT 3	0.0	0.0	0.0	Kidney cancer 1	0.0	0.0	0.0
Metastatic melanoma 1	15.8	24.1	15.2	Kidney NAT 1	0.1	0.0	0.1
Melanoma 2	0.1	0.0	0.0	Kidney cancer 2	3.3	8.8	2.1
Melanoma 3	0.0	0.0	0.0	Kidney NAT 2	0.3	0.0	0.5
Metastatic melanoma 4	**100.0**	**100.0**	**100.0**	Kidney cancer 3	0.3	0.0	0.2
Metastatic melanoma 5	34.9	63.3	24.0	Kidney NAT 3	0.1	0.0	0.1
Bladder cancer 1	0.0	0.0	0.0	Kidney cancer 4	0.5	0.3	0.5
Bladder NAT 1	0.0	0.0	0.0	Kidney NAT 4	0.2	0.0	0.3
Bladder cancer 2	0.1	0.0	0.1				

CNS_neurodegeneration_v1.0 Summary: Ag4300/Ag4301 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.4 Summary: Ag4300 Highest expression of this gene is detected in a breast cancer MCF-7 cell line (CT=25). This gene codes for Greb 1 protein. High expression of this gene is upregulated in response to estrogen in MCF-7 (Ghosh et al., 2000, Cancer Res 60(22):6367-75, PMID: 11103799). In addition, high to moderate levels of expression of this gene is also seen in number of cell lines derived from melanoma, ovarian, breast, lung, liver, renal, colon and brain cancers. Therefore, expression of this gene may be used as diagnostic marker for detection of these cancers. Furthermore, therapeutic modulation of this gene may be useful in the treatment of these cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, adrenal gland, pituitary gland, skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of

this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Interestingly, this gene is expressed at much higher levels in fetal (CT=29.6) when compared to adult liver (CT=35.9). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases.

High expression of this gene is also detected in adult lung (CT=26). Expression of this gene is higher in adult as compared to fetal lung (CT=31). Therefore, expression of this gene may be used to distinguish between adult and fetal lung.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 1.3D Summary: Ag3041/Ag3042 Two experiments with same probe and primer sets are in excellent agreement. Highest expression of this gene is detected in a breast cancer MCF-7 cell line (CTs=26.9). Moderate to low levels of expression of this gene is also seen in ovarian, breast, lung, liver, and brain cancer cell lines, brain and tissues with metabolic and endocrine function such as adipose, skeletal muscle, fetal heart, adrenal and pituatary glands. Please see panel 1.4 for further discussion on the utility of this gene.

Panel 2.2 Summary: Ag3041 Highest expression of this gene is detected in normal uterus (CT=30.9). Moderate to low levels of expression of this gene are also seen in both cancer and normal prostate, breast, and uterus. Therefore, therapeutic modulation of this gene may be useful in the treatment of these cancers.

Panel 3D Summary: Ag3041 Highest expression of this gene is detected in a neuroblastoma SK-N-SH cell line (CT=32.9). In addition, moderate to low levels of expression of this gene is also seen in cancer cell line derived from small lung cancer, B

and T cell lymphoma, and Wilm's tumor. Ag4300 Highest expression of this gene is seen in small lung cancer and melanoma cell line (CT=31.7).

Therefore, therapeutic modulation of this gene may be useful in the treatment of neuroblastoma, small lung cancer, B and T cell lymphoma and Wilm's tumor.

Panel 4.1D Summary: Ag4301 Highest expression of this gene is detected in eosinophils (CT=30.7). Differential gene expression is observed in the eosinophil cell line EOL-1 under resting conditions over that in EOL-1 cells stimulated by phorbol ester and ionomycin (CT=39). Thus, this gene may be involved in eosinophil function. Antibodies raised against this protein that stimulate its activity may be useful in the reduction of eosinophil activation and in the treatment of asthma and allergy and T cell-mediated autoimmune and inflammatory diseases.

Moderate levels of expression of this gene are also detected in kidney. Therefore, therapeutic modulation of this gene may be useful in kidney related diseases including lupus and glomerulonephritis.

Ag4300 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3041/Ag3042 Results from two experiments with this gene are not included. The amp plot indicates that there were experimental difficulties with this run. (Data not shown).

general oncology screening panel_v_2.4 Summary: Ag3042/Ag4300/Ag4301 Three experiments with different probe and primer sets are in excellent agreement. Highest expression of this gene is detected in metastatic melanoma (CTs=25-25.9). In addition, moderate to high expression of this gene is also detected in lung, prostate and kidney cancers. Thus, expression of this gene may be used as diagnostic marker for the detection of metastic melanoma, lung, prostate and kidney cancers.

B. CG106417-01: von Willebrand factor like protein

Expression of gene CG106417-01 was assessed using the primer-probe set Ag4470, described in Table BA. Results of the RTQ-PCR runs are shown in Tables BB, BC, BD, BE and BF.

Table BA. Probe Name Ag4470

Primers		Length	Start Position	SEQ ID No
Forward	5'-gcatcaggtgtacagacattga-3'	22	441	227
Probe	TET-5'-cgaatgtgtaacctcctcctgcgag-3'-TAMRA	25	463	228
Reverse	5'-acaaacccaccttctgtgttc-3'	21	499	229

Table BB. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag4470, Run 249008358	issue Name	Rel. Exp.(%) Ag4470, Run 249008358
110967 COPD-F	3.0	112427 Match Control Psoriasis-F	25.7
110980 COPD-F	8.7	112418 Psoriasis-M	4.3
110968 COPD-M	3.4	112723 Match Control Psoriasis-M	27.5
110977 COPD-M	38.2	112419 Psoriasis-M	2.6
110989 Emphysema-F	31.4	112424 Match Control Psoriasis-M	4.0
110992 Emphysema-F	3.3	112420 Psoriasis-M	13.7
110993 Emphysema-F	5.5	112425 Match Control Psoriasis-M	25.7
110994 Emphysema-F	2.1	104689 (MF) OA Bone-Backus	7.3
110995 Emphysema-F	15.4	104690 (MF) Adj "Normal" Bone-Backus	1.2
110996 Emphysema-F	2.0	104691 (MF) OA Synovium-Backus	11.3
110997 Asthma-M	0.8	104692 (BA) OA Cartilage-Backus	7.4
111001 Asthma-F	7.7	104694 (BA) OA Bone-Backus	2.0
111002 Asthma-F	5.5	104695 (BA) Adj "Normal" Bone-Backus	5.3
111003 Atopic Asthma-F	6.0	104696 (BA) OA Synovium-Backus	6.3
111004 Atopic Asthma-F	12.4	104700 (SS) OA Bone-Backus	5.6
111005 Atopic Asthma-F	5.6	104701 (SS) Adj "Normal" Bone-Backus	5.8
111006 Atopic Asthma-F	1.4	104702 (SS) OA Synovium-Backus	15.1
111417 Allergy-M	3.5	117093 OA Cartilage Rep7	12.2
112347 Allergy-M	5.8	112672 OA Bone5	97.3
112349 Normal Lung-F	6.1	112673 OA Synovium5	46.0
112357 Normal Lung-F	**100.0**	112674 OA Synovial Fluid cells5	32.5
112354 Normal Lung-M	69.3	117100 OA Cartilage Rep14	0.0
112374 Crohns-F	9.4	112756 OA Bone9	14.8
112389 Match Control Crohns-F	7.1	112757 OA Synovium9	17.4
112375 Crohns-F	7.4	112758 OA Synovial Fluid Cells9	5.2
112732 Match Control Crohns-F	6.7	117125 RA Cartilage Rep2	7.9
112725 Crohns-M	5.8	113492 Bone2 RA	1.5

112387 Match Control Crohns-M	0.0	113493 Synovium2 RA	0.0
112378 Crohns-M	4.7	113494 Syn Fluid Cells RA	0.0
112390 Match Control Crohns-M	52.5	113499 Cartilage4 RA	2.0
112726 Crohns-M	7.9	113500 Bone4 RA	1.7
112731 Match Control Crohns-M	13.1	113501 Synovium4 RA	2.3
112380 Ulcer Col-F	13.2	113502 Syn Fluid Cells4 RA	0.7
112734 Match Control Ulcer Col-F	8.4	113495 Cartilage3 RA	1.2
112384 Ulcer Col-F	2.8	113496 Bone3 RA	2.3
112737 Match Control Ulcer Col-F	2.8	113497 Synovium3 RA	0.0
112386 Ulcer Col-F	0.0	113498 Syn Fluid Cells3 RA	0.8
112738 Match Control Ulcer Col-F	1.6	117106 Normal Cartilage Rep20	5.7
112381 Ulcer Col-M	9.4	113663 Bone3 Normal	0.9
112735 Match Control Ulcer Col-M	25.2	113664 Synovium3 Normal	1.6
112382 Ulcer Col-M	7.6	113665 Syn Fluid Cells3 Normal	3.3
112394 Match Control Ulcer Col-M	0.0	117107 Normal Cartilage Rep22	3.5
112383 Ulcer Col-M	6.6	113667 Bone4 Normal	8.7
112736 Match Control Ulcer Col-M	2.4	113668 Synovium4 Normal	12.8
112423 Psoriasis-F	4.5	113669 Syn Fluid Cells4 Normal	24.3

Table BC. CNS_neurodegeneration_v1.0

Tissue Name	**Rel. Exp.(%) Ag4470, Run 224535165**	**issue Name**	**Rel. Exp.(%) Ag4470, Run 224535165**
AD 1 Hippo	13.7	Control (Path) 3 Temporal Ctx	2.8
AD 2 Hippo	22.2	Control (Path) 4 Temporal Ctx	31.6
AD 3 Hippo	6.3	AD 1 Occipital Ctx	17.8
AD 4 Hippo	10.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	35.1	AD 3 Occipital Ctx	7.9
AD 6 Hippo	95.9	AD 4 Occipital Ctx	11.3
Control 2 Hippo	15.8	AD 5 Occipital Ctx	9.3
Control 4 Hippo	23.7	AD 6 Occipital Ctx	20.3
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	5.8
AD 1 Temporal Ctx	15.0	Control 2 Occipital Ctx	36.3
AD 2 Temporal Ctx	14.8	Control 3 Occipital Ctx	9.4

AD 3 Temporal Ctx	2.6	Control 4 Occipital Ctx	10.7
AD 4 Temporal Ctx	23.7	Control (Path) 1 Occipital Ctx	54.7
AD 5 Inf Temporal Ctx	38.4	Control (Path) 2 Occipital Ctx	10.0
AD 5 SupTemporal Ctx	29.7	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	85.3	Control (Path) 4 Occipital Ctx	18.3
AD 6 Sup Temporal Ctx	**100.0**	Control 1 Parietal Ctx	7.4
Control 1 Temporal Ctx	7.7	Control 2 Parietal Ctx	33.2
Control 2 Temporal Ctx	28.5	Control 3 Parietal Ctx	9.6
Control 3 Temporal Ctx	16.7	Control (Path) 1 Parietal Ctx	22.4
Control 4 Temporal Ctx	14.5	Control (Path) 2 Parietal Ctx	28.1
Control (Path) 1 Temporal Ctx	32.3	Control (Path) 3 Parietal Ctx	2.2
Control (Path) 2 Temporal Ctx	34.9	Control (Path) 4 Parietal Ctx	44.1

Table BD. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4470, Run 222655825	issue Name	Rel. Exp.(%) Ag4470, Run 222655825
Adipose	4.8	Renal ca. TK-10	54.0
Melanoma* Hs688(A).T	3.3	Bladder	2.9
Melanoma* Hs688(B).T	3.1	Gastric ca. (liver met.) NCI-N87	2.3
Melanoma* M14	2.8	Gastric ca. KATO III	0.8
Melanoma* LOXIMVI	0.2	Colon ca. SW-948	0.5
Melanoma* SK-MEL-5	0.8	Colon ca. SW480	3.3
Squamous cell carcinoma SCC-4	0.6	Colon ca.* (SW480 met) SW620	16.2
Testis Pool	5.5	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	3.0	Colon ca. HCT-116	4.4
Prostate Pool	1.1	Colon ca. CaCo-2	94.0
Placenta	10.0	Colon cancer tissue	16.5
Uterus Pool	2.3	Colon ca. SW1116	0.6
Ovarian ca. OVCAR-3	0.8	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.4	Colon ca. SW-48	0.2
Ovarian ca. OVCAR-4	0.3	Colon Pool	2.6
Ovarian ca. OVCAR-5	1.6	Small Intestine Pool	10.8
Ovarian ca. IGROV-1	0.5	Stomach Pool	2.4
Ovarian ca. OVCAR-8	0.9	Bone Marrow Pool	1.0
Ovary	7.7	Fetal Heart	2.6
Breast ca. MCF-7	0.9	Heart Pool	1.7
Breast ca. MDA-MB-231	1.2	Lymph Node Pool	2.7
Breast ca. BT 549	1.8	Fetal Skeletal Muscle	2.3
Breast ca. T47D	4.9	Skeletal Muscle Pool	0.8
Breast ca. MDA-N	0.3	Spleen Pool	0.6

Breast Pool	2.4	Thymus Pool	16.3
Trachea	4.5	CNS cancer (glio/astro) U87-MG	5.7
Lung	7.9	CNS cancer (glio/astro) U-118-MG	2.7
Fetal Lung	3.8	CNS cancer (neuro;met) SK-N-AS	4.8
Lung ca. NCI-N417	3.9	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.9	CNS cancer (astro) SNB-75	5.2
Lung ca. NCI-H146	0.8	CNS cancer (glio) SNB-19	0.5
Lung ca. SHP-77	2.3	CNS cancer (glio) SF-295	8.3
Lung ca. A549	0.9	Brain (Amygdala) Pool	2.9
Lung ca. NCI-H526	2.9	Brain (cerebellum)	5.9
Lung ca. NCI-H23	1.4	Brain (fetal)	25.3
Lung ca. NCI-H460	2.2	Brain (Hippocampus) Pool	3.7
Lung ca. HOP-62	2.0	Cerebral Cortex Pool	4.6
Lung ca. NCI-H522	31.6	Brain (Substantia nigra) Pool	4.7
Liver	20.7	Brain (Thalamus) Pool	3.8
Fetal Liver	63.7	Brain (whole)	9.2
Liver ca. HepG2	**100.0**	Spinal Cord Pool	3.6
Kidney Pool	11.2	Adrenal Gland	4.2
Fetal Kidney	5.3	Pituitary gland Pool	0.8
Renal ca. 786-0	1.6	Salivary Gland	1.0
Renal ca. A498	0.8	Thyroid (female)	2.0
Renal ca. ACHN	2.2	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	12.9	Pancreas Pool	3.0

Table BE. Panel 4.1D

Tissue Name	**Rel. Ep.(%) Ag4470, Run 191882058**	**Tissue Name**	**Rel. Exp.(%) Ag4470, Run 191882058**
Secondary Th1 act	21.8	HUVEC IL-1beta	11.1
Secondary Th2 act	14.9	HUVEC IFN gamma	29.9
Secondary Tr1 act	11.3	HUVEC TNF alpha + IFN gamma	4.5
Secondary Th1 rest	5.3	HUVEC TNF alpha + IL4	45.7
Secondary Th2 rest	1.8	HUVEC IL-11	28.3
Secondary Tr1 rest	2.3	Lung Microvascular EC none	71.2
Primary Th1 act	42.0	Lung Microvascular EC TNFalpha + IL-1beta	27.7
Primary Th2 act	37.6	Microvascular Dermal EC none	38.4
Primary Tr1 act	42.3	Microsvasular Dermal EC TNFalpha + IL-1beta	24.1
Primary Th1 rest	1.1	Bronchial epithelium TNFalpha + IL1beta	5.0
Primary Th2 rest	1.3	Small airway epithelium none	6.6

Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	1.3
CD45RA CD4 lymphocyte act	5.9	Coronery artery SMC rest	10.3
CD45RO CD4 lymphocyte act	9.9	Coronery artery SMC TNFalpha + IL-1beta	1.8
CD8 lymphocyte act	19.2	Astrocytes rest	1.4
Secondary CD8 lymphocyte rest	10.4	Astrocytes TNFalpha + IL-1beta	3.1
Secondary CD8 lymphocyte act	4.5	KU-812 (Basophil) rest	29.5
CD4 lymphocyte none	0.6	KU-812 (Basophil) PMA/ionomycin	18.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	4.9	CCD1106 (Keratinocytes) none	2.3
LAK cells rest	1.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	3.5	Liver cirrhosis	10.2
LAK cells IL-2+IL-12	1.4	NCI-H292 none	16.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	9.0
LAK cells IL-2+ IL-18	2.3	NCI-H292 IL-9	32.5
LAK cells PMA/ionomycin	3.3	NCI-H292 IL-13	5.3
NK Cells IL-2 rest	3.9	NCI-H292 IFN gamma	15.5
Two Way MLR 3 day	4.8	HPAEC none	37.9
Two Way MLR 5 day	9.3	HPAEC TNF alpha + IL-1 beta	17.4
Two Way MLR 7 day	9.4	Lung fibroblast none	22.7
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	11.7
PBMC PWM	20.6	Lung fibroblast IL-4	17.7
PBMC PHA-L	18.3	Lung fibroblast IL-9	36.1
Ramos (B cell) none	4.5	Lung fibroblast IL-13	36.1
Ramos (B cell) ionomycin	9.2	Lung fibroblast IFN gamma	11.7
B lymphocytes PWM	20.3	Dermal fibroblast CCD1070 rest	1.3
B lymphocytes CD40L and IL-4	10.4	Dermal fibroblast CCD1070 TNF alpha	0.8
EOL-1 dbcAMP	1.9	Dermal fibroblast CCD1070 IL-1 beta	1.6
EOL-1 dbcAMP PMA/ionomycin	2.7	Dermal fibroblast IFN gamma	5.4
Dendritic cells none	5.1	Dermal fibroblast IL-4	**100.0**
Dendritic cells LPS	6.7	Dermal Fibroblasts rest	39.5
Dendritic cells anti-CD40	7.9	Neutrophils TNFa+LPS	0.0
Monocytes rest	1.0	Neutrophils rest	0.5
Monocytes LPS	1.6	Colon	0.5
Macrophages rest	13.0	Lung	0.7
Macrophages LPS	0.0	Thymus	59.9
HUVEC none	18.3	Kidney	0.7
HUVEC starved	13.5		

Table BF. general oncology screening panel_v 2.4

Tissue Name	Rel. Exp.(%) Ag4470, Run 260280484	Tissue ame	Rel. Exp.(%) Ag4470, Run 260280484
Colon cancer 1	1.0	Bladder NAT 2	0.1
Colon NAT 1	0.3	Bladder NAT 3	0.0
Colon cancer 2	0.0	Bladder NAT 4	1.1
Colon NAT 2	0.3	Prostate adenocarcinoma 1	4.3
Colon cancer 3	1.1	Prostate adenocarcinoma 2	1.5
Colon NAT 3	0.0	Prostate adenocarcinoma 3	1.8
Colon malignant cancer 4	2.2	Prostate adenocarcinoma 4	4.4
Colon NAT 4	0.0	Prostate NAT 5	1.0
Lung cancer 1	0.4	Prostate adenocarcinoma 6	0.5
Lung NAT 1	0.2	Prostate adenocarcinoma 7	0.2
Lung cancer 2	58.2	Prostate adenocarcinoma 8	0.7
Lung NAT 2	0.0	Prostate adenocarcinoma 9	1.7
Squamous cell carcinoma 3	1.3	Prostate NAT 10	0.6
Lung NAT 3	46.3	Kidney cancer 1	9.5
Metastatic melanoma 1	28.9	Kidney NAT 1	3.7
Melanoma 2	1.4	Kidney cancer 2	100.0
Melanoma 3	0.3	Kidney NAT 2	2.2
Metastatic melanoma 4	26.2	Kidney cancer 3	71.7
Metastatic melanoma 5	16.3	Kidney NAT 3	1.9
Bladder cancer 1	0.3	Kidney cancer 4	75.8
Bladder NAT 1	0.0	Kidney NAT 4	0.9
Bladder cancer 2	1.0		

AI_comprehensive panel_v1.0 Summary: Ag4470 These results confirm expression of this gene in cells involved in the immune response. Highest expression of this gene is seen in normal lung (CT=30.5). Please see Panel 4D for discussion of utility of this gene in inflammation.

CNS_neurodegeneration_v1.0 Summary: Ag4470 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at low but significant levels in the brain. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

General_screening_panel_v1.4 Summary: Ag4470 Highest expression of this gene is seen in a liver cancer cell line (CT=30), with moderate levels of expression seen in fetal and adult liver, and cell lines derived from colon, renal and lung cancers. Thus, expression of this gene could be used to differentiate liver derived tissue from other samples on this panel.

Panel 4.1D Summary: Ag4470 Highest expression of this gene in this experiment is detected in IL-4 treated dermal fibroblasts (CTs=30). In addition, this experiment shows low but significant levels of expresion in resting neutrophils (CT=33.2). In addition, this gene is expressed at moderate levels in IFN gamma stimulated dermal fibroblasts, activated lung fibroblasts, HPAECs, lung and dermal microvasculature, activated small airway and bronchial epithelium, activated NCI-H292 cells, acutely activated T cells, and activated B cells. Based on these levels of expression in T cells, activated B cells and cells in lung and skin, therapeutics that block the function of this gene product may be useful as therapeutics that reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases in which activated B cells present antigens in the generation of the aberrant immune response and in treating T-cell mediated diseases, including Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, allergy, emphysema, rheumatoid arthritis, or psoriasis.

general oncology screening panel_v_2.4 Summary: Ag4470 Highest expression of this gene is seen in kidney cancer (CT=30). In addition, this gene is more highly expressed in lung and kidney cancer than in the corresponding normal adjacent tissue. Thus, expression of this gene could be used as a marker of these cancers. Furthemore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of lung and kidney cancer.

C. CG106417-04: von Willebrand factor like protein

Expression of gene CG106417-04 was assessed using the primer-probe sets Ag1294b, Ag746, Ag905 and Ag4726, described in Tables CA, CB, CC and CD. Results of the RTQ-PCR runs are shown in Tables CE, CF, CG, CH, CI, CJ and CK.

Table CA. Probe Name Ag1294b

Primers		Length	Start Position	SEQ ID No
Forward	5'-cattggcagctacaagtgttc-3'	21	408	230

Probe	TET-5'-ctgtcgaactggcttccaccttcat-3'-TAMRA	25	429	231
Reverse	5'-cctccgacactcgtttacatc-3'	21	475	232

Table CB. Probe Name Ag746

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-gcattggcagctacaagtgt-3'	20	407	233
Probe	TET-5'-ctgtcgaactggcttccaccttcat-3'-TAMRA	25	429	234
Reverse	5'-cctccgacactcgtttacatc-3'	21	475	235

Table CC. Probe Name Ag905

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-cattggcagctacaagtgttc-3'	21	408	236
Probe	TET-5'-ctgtcgaactggcttccaccttcat-3'-TAMRA	25	429	237
Reverse	5'-cctccgacactcgtttacatc-3'	21	475	238

Table CD. Probe Name Ag4726

Primers		Length	Start Position	SEQ ID No
Forward	5'-gtgtctgtctggctggaaac-3'	20	1226	239
Probe	TET-5'-tgcatctctcctgagtgtccttctgg-3'-TAMRA	26	1252	240
Reverse	5'-acaagtacagcaatccgtctgt-3'	22	1296	241

Table CE. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag1294b, Run 249007981	issue Name	Rel. Exp.(%) Ag1294b, Run 249007981
110967 COPD-F	6.6	112427 Match Control Psoriasis-F	30.8
110980 COPD-F	16.6	112418 Psoriasis-M	4.6
110968 COPD-M	3.9	112723 Match Control Psoriasis-M	23.8
110977 COPD-M	31.6	112419 Psoriasis-M	2.7
110989 Emphysema-F	45.1	112424 Match Control Psoriasis-M	1.9
110992 Emphysema-F	7.2	112420 Psoriasis-M	4.9
110993 Emphysema-F	5.8	112425 Match Control Psoriasis-M	25.9
110994 Emphysema-F	3.3	104689 (MF) OA Bone-Backus	12.9

110995 Emphysema-F	2.0	104690 (MF) Adj "Normal" Bone-Backus	3.7
110996 Emphysema-F	3.1	104691 (MF) OA Synovium-Backus	6.9
110997 Asthma-M	3.7	104692 (BA) OA Cartilage-Backus	21.3
111001 Asthma-F	2.8	104694 (BA) OA Bone-Backus	6.6
111002 Asthma-F	5.3	104695 (BA) Adj "Normal" Bone-Backus	2.3
111003 Atopic Asthma-F	6.1	104696 (BA) OA Synovium-Backus	5.7
111004 Atopic Asthma-F	3.4	104700 (SS) OA Bone-Backus	6.2
111005 Atopic Asthma-F	3.9	104701 (SS) Adj "Normal" Bone-Backus	3.8
111006 Atopic Asthma-F	2.4	104702 (SS) OA Synovium-Backus	15.4
111417 Allergy-M	6.6	117093 OA Cartilage Rep7	18.0
112347 Allergy-M	3.3	112672 OA Bone5	90.1
112349 Normal Lung-F	3.2	112673 OA Synovium5	63.7
112357 Normal Lung-F	**100.0**	112674 OA Synovial Fluid cells5	32.3
112354 Normal Lung-M	58.6	117100 OA Cartilage Rep14	3.3
112374 Crohns-F	7.5	112756 OA Bone9	7.0
112389 Match Control Crohns-F	3.5	112757 OA Synovium9	12.2
112375 Crohns-F	5.1	112758 OA Synovial Fluid Cells9	3.9
112732 Match Control Crohns-F	0.5	117125 RA Cartilage Rep2	4.6
112725 Crohns-M	10.6	113492 Bone2 RA	2.4
112387 Match Control Crohns-M	3.5	113493 Synovium2 RA	1.1
112378 Crohns-M	1.7	113494 Syn Fluid Cells RA	1.4
112390 Match Control Crohns-M	55.5	113499 Cartilage4 RA	1.4
112726 Crohns-M	3.6	113500 Bone4 RA	0.5
112731 Match Control Crohns-M	13.9	113501 Synovium4 RA	1.7
112380 Ulcer Col-F	13.7	113502 Syn Fluid Cells4 RA	1.8
112734 Match Control Ulcer Col-F	5.6	113495 Cartilage3 RA	1.6
112384 Ulcer Col-F	3.9	113496 Bone3 RA	1.1
112737 Match Control Ulcer Col-F	3.3	113497 Synovium3 RA	0.0
112386 Ulcer Col-F	0.0	113498 Syn Fluid Cells3 RA	0.6
112738 Match Control Ulcer Col-F	0.0	117106 Normal Cartilage Rep20	4.5
112381 Ulcer Col-M	4.2	113663 Bone3 Normal	6.7
112735 Match Control Ulcer Col-M	18.2	113664 Synovium3 Normal	1.2
112382 Ulcer Col-M	4.2	113665 Syn Fluid Cells3 Normal	0.9
112394 Match Control Ulcer Col-M	0.0	117107 Normal Cartilage Rep22	1.3
112383 Ulcer Col-M	12.2	113667 Bone4 Normal	11.8

112736 Match Control Ulcer Col-M	2.0	113668 Synovium4 Normal	12.0
112423 Psoriasis-F	3.9	113669 Syn Fluid Cells4 Normal	10.7

Table CF. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag1294b, Run 206231468	Rel. Exp.(%) Ag4726, Run 224706360	issue Name	Rel. Exp.(%) Ag1294b, Run 206231468	Rel. Exp.(%) Ag4726, Run 224706360
AD 1 Hippo	11.2	11.6	Control (Path) 3 Temporal Ctx	1.5	11.4
AD 2 Hippo	22.5	23.5	Control (Path) 4 Temporal Ctx	19.2	20.3
AD 3 Hippo	4.7	0.0	AD 1 Occipital Ctx	15.8	17.4
AD 4 Hippo	8.7	15.2	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	37.6	35.6	AD 3 Occipital Ctx	1.2	3.6
AD 6 Hippo	**100.0**	**100.0**	AD 4 Occipital Ctx	17.8	7.9
Control 2 Hippo	28.7	21.9	AD 5 Occipital Ctx	8.7	17.6
Control 4 Hippo	30.4	40.3	AD 6 Occipital Ctx	12.3	30.8
Control (Path) 3 Hippo	6.9	3.6	Control 1 Occipital Ctx	0.0	3.0
AD 1 Temporal Ctx	16.3	26.1	Control 2 Occipital Ctx	27.4	34.6
AD 2 Temporal Ctx	31.6	25.2	Control 3 Occipital Ctx	5.4	2.8
AD 3 Temporal Ctx	3.8	5.6	Control 4 Occipital Ctx	6.7	15.4
AD 4 Temporal Ctx	10.9	36.1	Control (Path) 1 Occipital Ctx	56.3	85.3
AD 5 Inf Temporal Ctx	34.6	35.8	Control (Path) 2 Occipital Ctx	10.4	21.8
AD 5 SupTemporal Ctx	19.6	55.9	Control (Path) 3 Occipital Ctx	1.2	0.0
AD 6 Inf Temporal Ctx	73.7	76.8	Control (Path) 4 Occipital Ctx	6.3	5.0
AD 6 Sup Temporal Ctx	81.2	97.9	Control 1 Parietal Ctx	6.4	9.7
Control 1 Temporal Ctx	1.2	5.1	Control 2 Parietal Ctx	39.5	55.9
Control 2 Temporal Ctx	15.5	42.9	Control 3 Parietal Ctx	4.4	11.2
Control 3 Temporal Ctx	5.9	18.4	Control (Path) 1 Parietal Ctx	17.6	45.4
Control 4 Temporal Ctx	7.9	17.2	Control (Path) 2 Parietal Ctx	17.6	12.1

Control (Path) 1 Temporal Ctx	41.8	43.5	Control (Path) 3 Parietal Ctx	0.0	4.2
Control (Path) 2 Temporal Ctx	26.2	36.6	Control (Path) 4 Parietal Ctx	26.4	30.1

Table CG. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4726, Run 222842378	issue Name	Rel. Exp.(%) Ag4726, Run 222842378
Adipose	3.3	Renal ca. TK-10	41.8
Melanoma* Hs688(A).T	2.7	Bladder	1.8
Melanoma* Hs688(B).T	2.7	Gastric ca. (liver met.) NCI-N87	2.0
Melanoma* M14	4.8	Gastric ca. KATO III	0.6
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	0.6
Melanoma* SK-MEL-5	0.4	Colon ca. SW480	0.7
Squamous cell carcinoma SCC-4	0.2	Colon ca.* (SW480 met) SW620	12.8
Testis Pool	4.3	Colon ca. HT29	0.1
Prostate ca.* (bone met) PC-3	1.6	Colon ca. HCT-116	3.7
Prostate Pool	0.5	Colon ca. CaCo-2	31.9
Placenta	7.7	Colon cancer tissue	7.9
Uterus Pool	0.1	Colon ca. SW1116	1.0
Ovarian ca. OVCAR-3	0.7	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.6	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.3	Colon Pool	1.0
Ovarian ca. OVCAR-5	1.1	Small Intestine Pool	4.9
Ovarian ca. IGROV-1	1.4	Stomach Pool	3.4
Ovarian ca. OVCAR-8	0.7	Bone Marrow Pool	0.0
Ovary	5.0	Fetal Heart	0.7
Breast ca. MCF-7	0.4	Heart Pool	0.7
Breast ca. MDA-MB-231	0.5	Lymph Node Pool	2.6
Breast ca. BT 549	0.7	Fetal Skeletal Muscle	1.6
Breast ca. T47D	4.2	Skeletal Muscle Pool	1.0
Breast ca. MDA-N	0.2	Spleen Pool	0.4
Breast Pool	0.8	Thymus Pool	7.9
Trachea	1.3	CNS cancer (glio/astro) U87-MG	6.4
Lung	5.5	CNS cancer (glio/astro) U-118-MG	1.6
Fetal Lung	1.8	CNS cancer (neuro;met) SK-N-AS	4.4
Lung ca. NCI-N417	3.6	CNS cancer (astro) SF-539	0.2
Lung ca. LX-1	0.7	CNS cancer (astro) SNB-75	4.4
Lung ca. NCI-H146	0.8	CNS cancer (glio) SNB-19	1.1
Lung ca. SHP-77	0.3	CNS cancer (glio) SF-295	5.1
Lung ca. A549	0.8	Brain (Amygdala) Pool	2.5

Lung ca. NCI-H526	2.1	Brain (cerebellum)	7.3
Lung ca. NCI-H23	0.8	Brain (fetal)	12.2
Lung ca. NCI-H460	1.2	Brain (Hippocampus) Pool	1.9
Lung ca. HOP-62	0.5	Cerebral Cortex Pool	2.6
Lung ca. NCI-H522	20.2	Brain (Substantia nigra) Pool	2.1
Liver	11.6	Brain (Thalamus) Pool	3.9
Fetal Liver	61.1	Brain (whole)	8.5
Liver ca. HepG2	**100.0**	Spinal Cord Pool	1.9
Kidney Pool	6.7	Adrenal Gland	2.6
Fetal Kidney	2.0	Pituitary gland Pool	0.6
Renal ca. 786-0	1.7	Salivary Gland	0.9
Renal ca. A498	1.3	Thyroid (female)	1.8
Renal ca. ACHN	2.5	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	10.6	Pancreas Pool	1.1

Table CH. Panel 1.2

Tissue Name	**Rel. Exp.(% Ag746, Run 115163442**	**Rel. Exp.(%) Ag746, Run 119442272**	**Tissue Name**	**Rel. Exp.(%) Ag746, Run 115163442**	**Rel. Exp.(%) Ag746, Run 119442272**
Endothelial cells	12.3	5.9	Renal ca. 786-0	0.0	0.0
Heart (Fetal)	0.0	0.0	Renal ca. A498	0.0	0.0
Pancreas	0.0	0.0	Renal ca. RXF 393	0.0	0.0
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. ACHN	0.0	0.0
Adrenal Gland	0.0	0.2	Renal ca. UO-31	0.0	0.0
Thyroid	0.1	0.0	Renal ca. TK-10	0.0	0.0
Salivary gland	0.0	0.0	Liver	32.8	53.2
Pituitary gland	0.2	0.1	Liver (fetal)	72.7	**100.0**
Brain (fetal)	2.4	16.0	Liver ca. (hepatoblast) HepG2	**100.0**	94.0
Brain (whole)	0.0	0.3	Lung	0.0	0.0
Brain (amygdala)	0.0	0.0	Lung (fetal)	0.0	0.0
Brain (cerebellum)	0.0	0.0	Lung ca. (small cell) LX-1	0.0	0.0
Brain (hippocampus)	0.0	0.0	Lung ca. (small cell) NCI-H69	0.0	0.0
Brain (thalamus)	0.0	0.0	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Cerebral Cortex	0.0	0.0	Lung ca. (large cell)NCI-H460	0.0	0.0
Spinal cord	0.0	0.0	Lung ca. (non-sm. cell) A549	0.0	0.0

glio/astro U87-MG	0.0	0.0	Lung ca. (non-s.cell) NCI-H23	0.0	0.0
glio/astro U-118-MG	0.0	0.0	Lung ca. (non-s.cell) HOP-62	0.0	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cl) NCI-H522	63.7	90.1
neuro*; met SK-N-AS	0.0	0.2	Lung ca. (squam.) SW 900	0.0	0.0
astrocytoma SF-539	0.0	0.0	Lung ca. (squam.) NCI-H596	0.0	0.0
astrocytoma SNB-75	0.0	0.0	Mammary gland	0.7	3.6
glioma SNB-19	0.0	0.0	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma U251	0.0	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
glioma SF-295	0.0	0.0	Breast ca.* (pl. ef) T47D	0.0	0.0
Heart	0.0	0.0	Breast ca. BT-549	0.0	0.0
Skeletal Muscle	0.0	0.0	Breast ca. MDA-N	0.0	0.0
Bone marrow	0.0	0.0	Ovary	0.5	11.7
Thymus	1.2	2.8	Ovarian ca. OVCAR-3	0.0	0.0
Spleen	0.0	0.0	Ovarian ca. OVCAR-4	0.0	0.0
Lymph node	0.0	0.0	Ovarian ca. OVCAR-5	0.0	0.0
Colorectal Tissue	0.0	0.0	Ovarian ca. OVCAR-8	0.0	0.0
Stomach	0.0	0.0	Ovarian ca. IGROV-1	0.0	0.0
Small intestine	0.0	0.0	Ovarian ca. (ascites) SK-OV-3	0.0	0.0
Colon ca. SW480	0.0	0.0	Uterus	0.0	0.0
Colon ca.* SW620 (SW480 met)	1.1	1.9	Placenta	34.4	39.5
Colon ca. HT29	0.0	0.0	Prostate	0.0	0.0
Colon ca. HCT-116	0.0	0.0	Prostate ca.* (bone met) PC-3	0.0	0.0
Colon ca. CaCo-2	46.3	56.6	Testis	1.0	3.5
Colon ca. Tissue (ODO3866)	0.0	0.0	Melanoma Hs688(A).T	0.0	0.0
Colon ca. HCC-2998	0.0	0.0	Melanoma* (met) Hs688(B).T	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.0	0.0	Melanoma UACC-62	0.0	0.0
Bladder	0.0	0.0	Melanoma M14	0.0	0.0

Trachea	0.0	0.0	Melanoma LOX IMVI	0.0	0.0
Kidney	0.0	0.0	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney (fetal)	0.1	0.9			

Table CI. Panel 2D

Tissue Name	Rel. Exp.(% Ag746, Run 147127131	Rel. Exp.(%) Ag746, Run 148019631	Tissue Name	Rel. Exp.(%) Ag746, Run 147127131	Rel. Exp.(%) Ag746, Run 148019631
Normal Colon	18.3	21.8	Kidney Margin 8120608	6.5	6.4
CC Well to Mod Diff (ODO3866)	16.5	23.7	Kidney Cancer 8120613	2.2	0.7
CC Margin (ODO3866)	3.1	0.0	Kidney Margin 8120614	6.3	3.0
CC Gr.2 rectosigmoid (ODO3868)	0.0	0.8	Kidney Cancer 9010320	10.9	16.5
CC Margin (ODO3868)	0.5	2.0	Kidney Margin 9010321	9.0	11.3
CC Mod Diff (ODO3920)	1.2	2.3	Normal Uterus	4.3	6.3
CC Margin (ODO3920)	1.3	2.6	Uterus Cancer 064011	13.4	17.7
CC Gr.2 ascend colon (ODO3921)	3.4	4.4	Normal Thyroid	9.1	14.9
CC Margin (ODO3921)	1.3	0.0	Thyroid Cancer 064010	6.4	5.9
CC from Partial Hepatectomy (ODO4309) Mets	8.4	1.9	Thyroid Cancer A302152	4.4	5.1
Liver Margin (ODO4309)	49.7	41.5	Thyroid Margin A302153	12.0	22.1
Colon mets to lung (OD04451-01)	0.3	5.3	Normal Breast	9.9	14.3
Lung Margin (OD04451-02)	0.0	1.8	Breast Cancer (OD04566)	0.4	0.2
Normal Prostate 6546-1	9.1	12.1	Breast Cancer (OD04590-01)	5.3	3.9
Prostate Cancer (OD04410)	2.0	9.7	Breast Cancer Mets (OD04590-03)	4.0	10.4
Prostate Margin (OD04410)	16.8	20.3	Breast Cancer Metastasis (OD04655-05)	7.2	4.4
Prostate Cancer (OD04720-01)	13.5	14.4	Breast Cancer 064006	5.2	3.3

Prostate Margin (OD04720-02)	14.0	22.4	Breast Cancer 1024	12.1	18.6
Normal Lung 061010	6.8	11.7	Breast Cancer 9100266	2.7	5.3
Lung Met to Muscle (ODO4286)	1.8	0.7	Breast Margin 9100265	5.0	5.8
Muscle Margin (ODO4286)	11.5	13.1	Breast Cancer A209073	0.5	1.8
Lung Malignant Cancer (OD03126)	1.5	6.0	Breast Margin A209073	1.7	0.4
Lung Margin (OD03126)	4.8	2.4	Normal Liver	39.5	47.0
Lung Cancer (OD04404)	4.2	2.3	Liver Cancer 064003	4.2	0.6
Lung Margin (OD04404)	9.0	10.4	Liver Cancer 1025	66.4	74.2
Lung Cancer (OD04565)	0.3	0.0	Liver Cancer 1026	36.1	42.6
Lung Margin (OD04565)	0.4	0.3	Liver Cancer 6004-T	**100.0**	**100.0**
Lung Cancer (OD04237-01)	10.7	11.1	Liver Tissue 6004-N	22.8	34.4
Lung Margin (OD04237-02)	4.9	5.4	Liver Cancer 6005-T	39.2	35.4
Ocular Mel Met to Liver (ODO4310)	10.5	11.9	Liver Tissue 6005-N	33.2	38.2
Liver Margin (ODO4310)	22.4	32.8	Normal Bladder	6.6	4.9
Melanoma Mets to Lung (OD04321)	0.0	0.0	Bladder Cancer 1023	1.0	4.8
Lung Margin (OD04321)	0.6	0.0	Bladder Cancer A302173	2.6	0.7
Normal Kidney	5.3	5.3	Bladder Cancer (OD04718-01)	0.0	0.7
Kidney Ca, Nuclear grade 2 (OD04338)	39.8	43.8	Bladder Normal Adjacent (OD04718-03)	3.5	14.4
Kidney Margin (OD04338)	4.8	6.4	Normal Ovary	50.7	47.3
Kidney Ca Nuclear grade 1/2 (OD04339)	3.0	0.3	Ovarian Cancer 064008	10.2	7.4
Kidney Margin (OD04339)	5.4	10.0	Ovarian Cancer (OD04768-07)	73.7	80.7
Kidney Ca, Clear cell type (OD04340)	18.2	19.2	Ovary Margin (OD04768-08)	2.6	0.8
Kidney Margin (OD04340)	9.0	10.4	Normal Stomach	2.9	2.9

Kidney Ca, Nuclear grade 3 (OD04348)	5.2	8.3	Gastric Cancer 9060358	0.0	1.1
Kidney Margin (OD04348)	6.9	4.7	Stomach Margin 9060359	2.4	0.3
Kidney Cancer (OD04622-01)	41.8	45.4	Gastric Cancer 9060395	0.5	1.1
Kidney Margin (OD04622-03)	1.9	1.4	Stomach Margin 9060394	5.2	2.0
Kidney Cancer (OD04450-01)	9.2	6.2	Gastric Cancer 9060397	3.4	7.0
Kidney Margin (OD04450-03)	10.2	9.0	Stomach Margin 9060396	1.4	0.0
Kidney Cancer 8120607	2.2	1.7	Gastric Cancer 064005	1.3	6.0

Table CJ. Panel 4.1D

Tissue Name	Rel. Exp.(% Ag1294b, Run 200065765	Rel. Exp.(%) Ag4726, Run 204150067	Tissue Name	Rel. Exp.(%) Ag1294b, Run 200065765	Rel. Exp.(%) Ag4726, Run 204150067
Secondary Th1 act	15.3	8.4	HUVEC IL-1beta	5.6	10.2
Secondary Th2 act	7.2	0.4	HUVEC IFN gamma	21.9	13.3
Secondary Tr1 act	5.5	3.1	HUVEC TNF alpha + IFN gamma	3.5	1.1
Secondary Th1 rest	6.7	0.5	HUVEC TNF alpha + IL4	31.2	19.1
Secondary Th2 rest	1.0	2.6	HUVEC IL-11	17.7	20.7
Secondary Tr1 rest	1.3	0.5	Lung Microvascular EC none	65.1	61.6
Primary Th1 act	26.6	24.8	Lung Microvascular EC TNFalpha + IL-1beta	34.4	30.4
Primary Th2 act	34.2	19.8	Microvascular Dermal EC none	42.3	29.9
Primary Tr1 act	40.3	27.9	Microsvasular Dermal EC TNFalpha + IL-1beta	16.7	7.6
Primary Th1 rest	0.3	0.0	Bronchial epithelium TNFalpha + IL1beta	2.4	4.4
Primary Th2 rest	0.5	0.0	Small airway epithelium none	1.7	4.2
Primary Tr1 rest	0.0	1.1	Small airway epithelium TNFalpha + IL-1beta	2.5	2.4
CD45RA CD4 lymphocyte act	7.7	2.2	Coronery artery SMC rest	9.0	2.1

CD45RO CD4 lymphocyte act	10.9	16.5	Coronery artery SMC TNFalpha + IL-1beta	5.2	4.1
CD8 lymphocyte act	11.0	9.9	Astrocytes rest	2.1	0.8
Secondary CD8 lymphocyte rest	11.8	8.9	Astrocytes TNFalpha + IL-1beta	2.2	1.2
Secondary CD8 lymphocyte act	4.7	1.9	KU-812 (Basophil) rest	10.2	14.9
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	11.1	8.6
2ry Th1/Th2/Tr1_anti-CD95 CH11	1.7	2.5	CCD1106 (Keratinocytes) none	0.0	0.9
LAK cells rest	0.0	1.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.6	0.0
LAK cells IL-2	3.1	1.7	Liver cirrhosis	6.8	6.0
LAK cells IL-2+IL-12	2.9	1.1	NCI-H292 none	21.3	10.3
LAK cells IL-2+IFN gamma	0.5	1.3	NCI-H292 IL-4	11.5	7.3
LAK cells IL-2+ IL-18	0.5	1.1	NCI-H292 IL-9	13.8	17.4
LAK cells PMA/ionomycin	1.0	4.2	NCI-H292 IL-13	19.9	6.7
NK Cells IL-2 rest	1.4	2.0	NCI-H292 IFN gamma	7.3	13.8
Two Way MLR 3 day	3.1	1.8	HPAEC none	20.4	28.9
Two Way MLR 5 day	5.0	4.2	HPAEC TNF alpha + IL-1 beta	21.5	15.4
Two Way MLR 7 day	4.7	4.0	Lung fibroblast none	23.5	15.7
PBMC rest	0.6	0.0	Lung fibroblast TNF alpha + IL-1 beta	8.8	9.2
PBMC PWM	11.5	9.9	Lung fibroblast IL-4	21.2	24.7
PBMC PHA-L	7.2	14.1	Lung fibroblast IL-9	16.8	18.2
Ramos (B cell) none	1.8	2.0	Lung fibroblast IL-13	33.2	19.8
Ramos (B cell) ionomycin	3.4	2.7	Lung fibroblast IFN gamma	19.1	7.8
B lymphocytes PWM	20.2	17.6	Dermal fibroblast CCD1070 rest	2.9	0.1
B lymphocytes CD40L and IL-4	12.2	11.0	Dermal fibroblast CCD1070 TNF alpha	0.0	0.2
EOL-1 dbcAMP	1.5	3.2	Dermal fibroblast CCD1070 IL-1 beta	1.5	4.5
EOL-1 dbcAMP PMA/ionomycin	1.1	0.5	Dermal fibroblast IFN gamma	45.1	32.8
Dendritic cells none	8.5	4.0	Dermal fibroblast IL-4	**100.0**	**100.0**
Dendritic cells LPS	6.4	5.9	Dermal Fibroblasts rest	53.6	39.2

Dendritic cells anti-CD40	8.7	4.7	Neutrophils TNFa+LPS	1.5	0.6
Monocytes rest	0.0	0.0	Neutrophils rest	10.2	0.1
Monocytes LPS	1.1	2.2	Colon	1.5	1.6
Macrophages rest	8.8	4.8	Lung	1.7	1.3
Macrophages LPS	0.0	0.0	Thymus	40.1	25.0
HUVEC none	10.1	8.5	Kidney	1.5	0.0
HUVEC starved	7.6	11.4			

Table CK. Panel 4D

Tissue Name	Rel. Exp.(% Ag1294b, Run 138944262	Rel. Exp.(%) Ag1294b, Run 139408252	Tissue Name	Rel. Exp.(%) Ag1294b, Run 138944262	Rel. Exp.(%) Ag1294b, Run 139408252
Secondary Th1 act	10.9	7.7	HUVEC IL-1beta	4.1	1.7
Secondary Th2 act	6.4	8.0	HUVEC IFN gamma	21.0	13.7
Secondary Tr1 act	11.3	9.3	HUVEC TNF alpha + IFN gamma	2.8	0.6
Secondary Th1 rest	3.4	2.7	HUVEC TNF alpha + IL4	30.8	25.7
Secondary Th2 rest	1.5	2.5	HUVEC IL-11	11.6	7.3
Secondary Tr1 rest	1.4	2.0	Lung Microvascular EC none	24.1	20.0
Primary Th1 act	48.0	46.0	Lung Microvascular EC TNFalpha + IL-1beta	8.0	12.2
Primary Th2 act	38.7	27.7	Microvascular Dermal EC none	64.6	45.7
Primary Tr1 act	72.2	55.5	Microsvasular Dermal EC TNFalpha + IL-1beta	18.4	11.7
Primary Th1 rest	3.1	2.3	Bronchial epithelium TNFalpha + IL1beta	5.2	5.4
Primary Th2 rest	1.0	0.8	Small airway epithelium none	4.0	3.2
Primary Tr1 rest	1.1	0.5	Small airway epithelium TNFalpha + IL-1beta	8.2	4.5
CD45RA CD4 lymphocyte act	2.9	1.8	Coronery artery SMC rest	5.8	6.3
CD45RO CD4 lymphocyte act	18.6	12.2	Coronery artery SMC TNFalpha + IL-1beta	4.5	5.1
CD8 lymphocyte act	17.8	6.8	Astrocytes rest	0.8	0.5
Secondary CD8 lymphocyte rest	6.8	6.0	Astrocytes TNFalpha + IL-1beta	3.6	1.9

Secondary CD8 lymphocyte act	5.5	4.1	KU-812 (Basophil) rest	16.0	11.1
CD4 lymphocyte none	0.0	0.2	KU-812 (Basophil) PMA/ionomycin	12.3	9.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.9	3.1	CCD1106 (Keratinocytes) none	0.0	0.5
LAK cells rest	1.4	0.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.7	0.4
LAK cells IL-2	3.8	2.2	Liver cirrhosis	8.4	3.8
LAK cells IL-2+IL-12	3.0	0.8	Lupus kidney	2.0	3.2
LAK cells IL-2+IFN gamma	2.0	1.7	NCI-H292 none	21.9	25.7
LAK cells IL-2+ IL-18	0.5	0.2	NCI-H292 IL-4	15.7	12.3
LAK cells PMA/ionomycin	0.7	1.3	NCI-H292 IL-9	20.6	14.7
NK Cells IL-2 rest	0.7	0.7	NCI-H292 IL-13	8.3	5.7
Two Way MLR 3 day	1.1	2.5	NCI-H292 IFN gamma	5.1	8.2
Two Way MLR 5 day	2.5	2.8	HPAEC none	18.7	23.8
Two Way MLR 7 day	4.5	5.0	HPAEC TNF alpha + IL-1 beta	11.9	12.9
PBMC rest	0.0	0.0	Lung fibroblast none	15.7	13.5
PBMC PWM	41.8	29.1	Lung fibroblast TNF alpha + IL-1 beta	6.9	4.7
PBMC PHA-L	34.4	21.8	Lung fibroblast IL-4	25.0	16.6
Ramos (B cell) none	4.7	2.4	Lung fibroblast IL-9	14.7	15.8
Ramos (B cell) ionomycin	9.2	5.8	Lung fibroblast IL-13	40.3	32.5
B lymphocytes PWM	51.8	51.4	Lung fibroblast IFN gamma	15.4	17.4
B lymphocytes CD40L and IL-4	10.2	12.3	Dermal fibroblast CCD1070 rest	0.5	0.9
EOL-1 dbcAMP	0.3	0.2	Dermal fibroblast CCD1070 TNF alpha	0.9	0.8
EOL-1 dbcAMP PMA/ionomycin	0.4	1.8	Dermal fibroblast CCD1070 IL-1 beta	0.6	0.6
Dendritic cells none	6.7	3.8	Dermal fibroblast IFN gamma	32.1	18.4
Dendritic cells LPS	4.7	3.1	Dermal fibroblast IL-4	**100.0**	**100.0**
Dendritic cells anti-CD40	6.0	5.6	IBD Colitis 2	0.0	0.0
Monocytes rest	0.0	0.0	IBD Crohn's	0.3	0.8
Monocytes LPS	0.7	0.8	Colon	1.4	0.5
Macrophages rest	19.8	9.9	Lung	0.5	0.8

Macrophages LPS	0.7	0.5	Thymus	2.9	4.3
HUVEC none	9.3	10.2	Kidney	65.5	47.3
HUVEC starved	19.2	13.1			

AI_comprehensive panel_v1.0 Summary: Ag1294b Expression of this gene in this panel confirms expression of this gene in cells involved in the immune response. Highest expression of this gene is seen in normal lung (CT=30.5). Please see Panel 4D for discussion of utility of this gene in inflammation.

CNS_neurodegeneration_v1.0 Summary: Ag1294b/Ag4726 Two experiments with different probe and primer sets produce results that are in reasonable agreement. This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at low but significant levels in the brain. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

General_screening_panel_v1.4 Summary: Ag4726 Highest expression of this gene is seen in a liver cancer cell line (CTs=30), with moderate levels of expression seen in fetal and adult liver, and cell lines derived from colon, renal and lung cancers. Thus, expression of this gene could be used to differentiate liver derived tissue from other samples on this panel.

Panel 1.2 Summary: Ag746 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression of this gene in a liver cancer cell line (CTs=27). High levels of expression are also seen in fetal and adult liver tissue, a colon cancer cell line and a lung cancer cell line. Thus, expression of this gene could be used to differentiate liver derived samples, the colon cancer cell line and the lung cancer cell line from other samples on this panel. Expression of this gene could also be used as a diagnostic marker to detect the presence of colon and lung cancers.

Moderate expression is also seen in the fetal brain, placenta, and endothelial cells.

Panel 2D Summary: Ag746 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression of this gene in liver cancer (CTs=31). The prominent expression in liver derived tissue is consistent with the results in Panel 1.2. Moderate levels of expression are also evident in samples from ovarian cancer and kidney cancer. Furthermore, expression of this gene is higher in these cancers

than in the normal adjacent tissue. Thus, expression of this gene could be used to differentiate between liver derived samples and other samples on this panel and as a marker to detect the presence of liver, kidney, and ovarian cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of liver, kidney, and ovarian cancers.

Panel 4.1D Summary: Ag1294b/Ag4726 Results from three experiments with three different probe and primer sets are in agreement with the expression profile in Panel 4D, with highest expression of this gene in this experiment in IL-4 treated dermal fibroblasts (CTs=30). In addition, this experiment shows low but significant levels of expresion in resting neutrophils (CT=33.2), a sample absent in Panel 4D. Please see Panel 4D for discussion of utility of this gene in inflammation.

Panel 4D Summary: Ag1294b Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression of this gene in IL-4 treated dermal fibroblasts (CTs=30). In addition, this gene is expressed at moderate levels in IFN gamma stimulated dermal fibroblasts, activated lung fibroblasts, HPAECs, lung and dermal microvasculature, activated small airway and bronchial epithelium, activated NCI-H292 cells, acutely activated T cells, and activated B cells.

Based on these levels of expression in T cells, activated B cells and cells in lung and skin, therapeutics that block the function of this gene product may be useful as therapeutics that reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases in which activated B cells present antigens in the generation of the aberrant immune response and in treating T-cell mediated diseases, including Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, allergy, emphysema, rheumatoid arthritis, or psoriasis.

D. CG108901-03: CYTOKINE RECEPTOR

Expression of full length physical clone CG108901-03 was assessed using the primer-probe set Ag6889, described in Table DA. Results of the RTQ-PCR runs are shown in Table DB.

Table DA. Probe Name Ag6889

Primers		Length	Start Position	SEQ ID No

Forward	5'-aaggaaagggccctgcct-3'	18	61	242
Probe	TET-5'-caacgtccaccagctgcaccatcac-3'-TAMRA	25	90	243
Reverse	5'-gaaccatggagaacagctgga-3'	21	120	244

Table DB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag6889, Run 278388254	issue Name	Rel. Exp.(%) Ag6889, Run 278388254
Adipose	0.1	Renal ca. TK-10	1.0
Melanoma* Hs688(A).T	0.0	Bladder	0.6
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.4	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.3
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.1
Prostate Pool	0.1	Colon ca. CaCo-2	0.2
Placenta	**100.0**	Colon cancer tissue	0.2
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.1	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.1	Colon Pool	0.1
Ovarian ca. OVCAR-5	0.1	Small Intestine Pool	0 1
Ovarian ca. IGROV-1	0.4	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.3	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.1	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.4
Breast Pool	0.1	Thymus Pool	0.2
Trachea	0.1	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.1	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.2
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.1

Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.3
Lung ca. NCI-H23	0.8	Brain (fetal)	0.2
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.1
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.1
Lung ca. NCI-H522	0.4	Brain (Substantia nigra) Pool	0.1
Liver	0.1	Brain (Thalamus) Pool	0.1
Fetal Liver	0.3	Brain (whole)	0.1
Liver ca. HepG2	0.3	Spinal Cord Pool	0.2
Kidney Pool	0.1	Adrenal Gland	0.3
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.3	Salivary Gland	0.1
Renal ca. A498	0.2	Thyroid (female)	0.0
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	0.1
Renal ca. UO-31	0.5	Pancreas Pool	0.0

General_screening_panel_v1.6 Summary: Ag6889 High expression of this gene is restricted to placenta. Thus, expression of this gene may be used as a marker to distinguish placenta from other samples. This gene codes for a splice variant of EBV-induced gene 3 (EBI3), encodes a 34-kDa glycoprotein which lacks a membrane-anchoring motif and is secreted. EBI3 is shown to be expressed in vivo by scattered cells in interfollicular zones of tonsil tissue, by cells associated with sinusoids in perifollicular areas of spleen tissue, and at very high levels by placental syncytiotrophoblasts (Devergne et al., 1996, J. Virol. 70: 1143-1153, PMID:8551575). In addition, EBI3 levels are strongly up-regulated in sera from pregnant women and gradually increased with gestational age. EBI3 is an important immunomodulator in the fetal-maternal relationship, possibly involved in NK cell regulation (Devergne et al., 2001, Am J Pathol 2001 Nov;159(5):1763-76, PMID: 11696437). Thus, therapeutic modulation of this gene or EBI3 protein encoded by this gene may be useful in the treatment of placenta or pregnancy related diseases.

E. CG108901-04: CYTOKINE RECEPTOR

Expression of full length physical clone CG108901-04 was assessed using the primer-probe set Ag7033, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB and EC.

Table EA. Probe Name Ag7033

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-ctcccactgcacctgtagct-3'	20	254	245
Probe	TET-5'-taacagaccacatcatcaagcccgacc-3'-TAMRA	27	313	246
Reverse	5'-accagccccgtgccttt-3'	17	342	247

Table EB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag7033, Run 282263480	issue Name	Rel. Exp.(%) Ag7033, Run 282263480
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	**100.0**	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0

Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table EC. Panel 4.1D

Tissue Name	**Rel. Ep.(%) Ag7033, Run 312115300**	**Tissue Name**	**Rel. Exp.(%) Ag7033, Run 312115300**
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.1
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0

CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	**100.0**
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.1
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.1
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.2	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.3	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.1	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.3	Colon	0.0
Macrophages rest	0.6	Lung	0.0
Macrophages LPS	0.1	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

General_screening_panel_v1.6 Summary: Ag7033 Low expression of this gene is restricted to placenta. Thus, expression of this gene may be used as a marker to distinguish placenta from other samples. This gene codes for a splice variant of

EBV-induced gene 3 (EBI3), a 34-kDa glycoprotein that lacks a membrane-anchoring motif and is secreted. EBI3 is shown to be expressed in vivo by scattered cells in interfollicular zones of tonsil tissue, by cells associated with sinusoids in perifollicular areas of spleen tissue, and at very high levels by placental syncytiotrophoblasts (Devergne et al., 1996, J. Virol. 70: 1143-1153, PMID:8551575). In addition, EBI3 levels are strongly up-regulated in sera from pregnant women and gradually increased with gestational age. EBI3 is an important immunomodulator in the fetal-maternal relationship, possibly involved in NK cell regulation (Devergne et al., 2001, Am J Pathol 2001 Nov;159(5):1763-76, PMID: 11696437). Thus, therapeutic modulation of this gene or EBI3 protein encoded by this gene may be useful in the treatment of placenta or pregnancy related diseases.

Panel 4.1D Summary: Ag7033 High expression of this gene is restricted to PMA/ionomycin activated basophils (CT=27.9). Basophils release histamines and other biological modifiers in reponse to allergens and play an important role in the pathology of asthma and hypersensitivity reactions. Therefore, therapeutics designed against the putative protein encoded by this gene may reduce or inhibit inflammation by blocking basophil function in these diseases. In addition, these cells are a reasonable model for the inflammatory cells that take part in various inflammatory lung and bowel diseases, such as asthma, Crohn's disease, and ulcerative colitis. Therefore, therapeutics that modulate the function of this gene product may reduce or eliminate the symptoms of patients suffering from asthma, Crohn's disease, and ulcerative colitis.

F. CG126129-02: Epithelium differentiation factor (PEDF) (Similar to Serine or Cysteine proteinase inhibitor)

Expression of full length physical clone CG126129-02 was assessed using the primer-probe set Ag7039, described in Table FA.

Table FA. Probe Name Ag7039

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-ggtggaggaggaggatcct-3'	19	169	248
Probe	TET-5'-cttcaaagtccccgtgaacaagctgg-3'-TAMRA	26	190	249
Reverse	5'-tggattctgttcgctggat-3'	19	259	250

General_screening_panel_v1.6 Summary: Ag7039 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

G. CG142202-03: CRL-2

Expression of full length physical clone CG142202-03 was assessed using the primer-probe set Ag4530, described in Table GA. Results of the RTQ-PCR runs are shown in Tables GB and GC.

Table GA. Probe Name Ag4530

Primers		Length	Start Position	SEQ ID No
Forward	5'-acatggaatgccagcaaatac-3'	21	994	251
Probe	TET-5'-tccaggaccaacctgactttccacta-3'-TAMRA	26	968	252
Reverse	5'-actggtcataggcctcatcac-3'	21	936	253

Table GB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4530, Run 222735181	issue Name	Rel. Exp.(%) Ag4530, Run 222735181
Adipose	2.3	Renal ca. TK-10	19.1
Melanoma* Hs688(A).T	0.0	Bladder	0.6
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	3.1
Melanoma* M14	0.0	Gastric ca. KATO III	0.5
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	6.3	Colon ca. SW480	1.3
Squamous cell carcinoma SCC-4	0.3	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.5	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	1.2
Prostate Pool	0.5	Colon ca. CaCo-2	0.0
Placenta	0.2	Colon cancer tissue	1.5
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.1
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	1.6	Colon Pool	3.1
Ovarian ca. OVCAR-5	1.4	Small Intestine Pool	0.7
Ovarian ca. IGROV-1	0.3	Stomach Pool	0.5
Ovarian ca. OVCAR-8	0.8	Bone Marrow Pool	2.3
Ovary	0.6	Fetal Heart	0.0

Breast ca. MCF-7	0.1	Heart Pool	0.7
Breast ca. MDA-MB-231	66.0	Lymph Node Pool	2.9
Breast ca. BT 549	13.2	Fetal Skeletal Muscle	0.3
Breast ca. T47D	1.7	Skeletal Muscle Pool	0.5
Breast ca. MDA-N	0.9	Spleen Pool	1.1
Breast Pool	2.0	Thymus Pool	2.5
Trachea	4.2	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	1.9
Fetal Lung	0.4	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.3
Lung ca. LX-1	0.3	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.2
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	2.2	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.5	Brain (fetal)	0.0
Lung ca. NCI-H460	**100.0**	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	1.3	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.8	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	1.9	Adrenal Gland	0.0
Fetal Kidney	0.2	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.7
Renal ca. ACHN	1.0	Pancreatic ca. CAPAN2	5.1
Renal ca. UO-31	0.0	Pancreas Pool	1.1

Table GC. Panel 4.1D

Tissue Name	**Rel. Exp.() Ag4530, Run 198383582**	**Tissue Name**	**Rel. Exp.(%) Ag4530, Run 198383582**
Secondary Th1 act	1.8	HUVEC IL-1beta	0.0
Secondary Th2 act	20.9	HUVEC IFN gamma	0.0
Secondary Tr1 act	15.5	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	1.3	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	3.0	HUVEC IL-11	0.1
Secondary Tr1 rest	5.2	Lung Microvascular EC none	0.0
Primary Th1 act	0.5	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	4.0	Microvascular Dermal EC none	0.0

Primary Tr1 act	0.8	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	3.4	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	3.7	Small airway epithelium none	0.0
Primary Tr1 rest	4.2	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.5	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.7	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.8	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.5	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	1.7	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.5	KU-812 (Basophil) PMA/ionomycin	0.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	19.1	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.7	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.2	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.5	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.3	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	6.2	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.9	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	1.3	HPAEC none	0.0
Two Way MLR 5 day	0.4	HPAEC TNF alpha + IL-1 beta	0.1
Two Way MLR 7 day	0.5	Lung fibroblast none	0.0
PBMC rest	0.1	Lung fibroblast TNF alpha + IL-1 beta	0.1
PBMC PWM	0.2	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.5	Lung fibroblast IL-9	0.0
Ramos (B cell) none	7.6	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	3.8	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.2	Dermal fibroblast CCD1070 rest	0.1
B lymphocytes CD40L and IL-4	0.6	Dermal fibroblast CCD1070 TNF alpha	1.4
EOL-1 dbcAMP	0.4	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	**100.0**	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.1	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	3.9	Dermal Fibroblasts rest	0.1
Dendritic cells anti-CD40	0.6	Neutrophils TNFa+LPS	0.8
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	9.0	Colon	0.1

Macrophages rest	0.2	Lung	0.0
Macrophages LPS	2.4	Thymus	0.4
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag4530 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag4530 Highest expression of this gene is detected in a lung cancer NCI-H460 cell line (CT=27.7). In addition, moderate levels of expression of this gene is also seen in cancer cell lines derived from melanoma, breast, pancreatic, lung, renal, brain and colon cancers. Thus, expression of this gene may be used as diagnostic marker to detect the presence of these cancers. Furthermore, therapeutic modulation of this gene may be useful in the treatment of melanoma, lung, breast, colon, renal, pancreatic and brain cancers.

Among the tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adipose, thyroid, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Interestingly, this gene is expressed at much higher levels in fetal (CT=34.7) when compared to adult liver (CT=40). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases.

Panel 4.1D Summary: Ag4530 Highest expression of this gene is detected in PMA/ionomycin treated eosinophils (CT=26.4). Expression of this gene is higher in activated as compared to resting eosinophil (CT=34.3). Thus, expression of this gene may be used to distinguish between resting and activated eosinophils and also from other samples used in this panel. In addition, expression of this gene in activated eosinophil suggests a role for this gene in eosinophil functions. Therefore, therapeutic modulation of this gene through the use of antibodies or small molecule drug may be useful in the

treatment of T cell-mediated autoimmune and inflammatory diseases including asthma and allergy and also hematopoietic disorders involving eosinphils, parasitic infections.

In addition, low to moderate levels of expression of this gene is also detected in T lymphocytes prepared under a number of conditions, as well as, in different activated cell types involved in inflammatory and autoimmune disorders such as dendritic cells, monocytes, macrophages, neutrophils and dermal fibroblasts. Dendritic cells and macrophages are powerful antigen-presenting cells (APC) whose function is pivotal in the initiation and maintenance of normal immune responses. Autoimmunity and inflammation may also be reduced by suppression of this function. Therefore, small molecule drugs and antibodies that antagonzie the function of this gene product may reduce or eliminate the symptoms in patients with several types of autoimmune and inflammatory diseases, such as lupus erythematosus, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or psoriasis.

H. CG142621-01: Hypothetical Membrane Protein

Expression of gene CG142621-01 was assessed using the primer-probe set Ag7570, described in Table HA.

Table HA. Probe Name Ag7570

Primers	Sequencs	Length	Start Position	SEQ ID No
Forward	5'-gccagcatccaactcagattat-3'	22	234	254
Probe	TET-5'-cacaatctccttacattgacagttttg atg-3'-TAMRA	30	260	255
Reverse	5'-ggattccaagttcttctagcaaa-3'	23	300	256

CNS_neurodegeneration_v1.0 Summary: Ag7570 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag7570 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

I. CG142761-01: Similar to histocompatibility 13

Expression of gene CG142761-01 was assessed using the primer-probe set Ag7623, described in Table IA. Results of the RTQ-PCR runs are shown in Tables IB and IC.

Table IA. Probe Name Ag7623

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-cccagcgccatgtaatg-3'	17	1293	257
Probe	TET-5'-atttgactcctcataacttgggcccc -3'-TAMRA	26	1350	258
Reverse	5'-gccgctggatccttagg-3'	17	1376	259

Table IB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7623, Run 311288617	issue Name	Rel. Exp.(%) Ag7623, Run 311288617
AD 1 Hippo	8.1	Control (Path) 3 Temporal Ctx	7.7
AD 2 Hippo	28.9	Control (Path) 4 Temporal Ctx	41.8
AD 3 Hippo	18.7	AD 1 Occipital Ctx	24.1
AD 4 Hippo	16.8	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	57.8	AD 3 Occipital Ctx	9.8
AD 6 Hippo	31.6	AD 4 Occipital Ctx	63.7
Control 2 Hippo	61.6	AD 5 Occipital Ctx	93.3
Control 4 Hippo	6.0	AD 6 Occipital Ctx	12.8
Control (Path) 3 Hippo	9.2	Control 1 Occipital Ctx	6.9
AD 1 Temporal Ctx	23.7	Control 2 Occipital Ctx	86.5
AD 2 Temporal Ctx	37.1	Control 3 Occipital Ctx	37.4
AD 3 Temporal Ctx	18.3	Control 4 Occipital Ctx	11.7
AD 4 Temporal Ctx	36.1	Control (Path) 1 Occipital Ctx	66.9
AD 5 Inf Temporal Ctx	72.2	Control (Path) 2 Occipital Ctx	21.6
AD 5 Sup Temporal Ctx	25.9	Control (Path) 3 Occipital Ctx	15.3
AD 6 Inf Temporal Ctx	16.7	Control (Path) 4 Occipital Ctx	17.7
AD 6 Sup Temporal Ctx	31.4	Control 1 Parietal Ctx	21.2
Control 1 Temporal Ctx	23.2	Control 2 Parietal Ctx	25.9
Control 2 Temporal Ctx	76.8	Control 3 Parietal Ctx	23.5
Control 3 Temporal Ctx	43.5	Control (Path) 1 Parietal Ctx	94.0
Control 3 Temporal Ctx	13.9	Control (Path) 2 Parietal Ctx	37.9
Control (Path) 1 Temporal Ctx	**100.0**	Control (Path) 3 Parietal Ctx	4.9
Control (Path) 2 Temporal Ctx	37.1	Control (Path) 4 Parietal Ctx	52.9

Table IC. Panel 4.1D

Tissue Name	Rel. Exp. %) Ag7623, Run 311288446	Tissue Name	Rel. Exp.(%) Ag7623, Run 311288446
Secondary Th1 act	9.3	HUVEC IL-1beta	9.9
Secondary Th2 act	2.4	HUVEC IFN gamma	3.7
Secondary Tr1 act	6.9	HUVEC TNF alpha + IFN gamma	9.9
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	8.2
Secondary Th2 rest	0.0	HUVEC IL-11	2.5
Secondary Tr1 rest	1.8	Lung Microvascular EC none	15.0
Primary Th1 act	2.6	Lung Microvascular EC TNFalpha + IL-1beta	6.0
Primary Th2 act	6.2	Microvascular Dermal EC none	6.6
Primary Tr1 act	8.9	Microsvasular Dermal EC TNFalpha + IL-1beta	20.3
Primary Th1 rest	2.0	Bronchial epithelium TNFalpha + IL1beta	10.6
Primary Th2 rest	0.0	Small airway epithelium none	9.6
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	11.5
CD45RA CD4 lymphocyte act	9.5	Coronery artery SMC rest	23.2
CD45RO CD4 lymphocyte act	2.5	Coronery artery SMC TNFalpha + IL-1beta	12.0
CD8 lymphocyte act	4.0	Astrocytes rest	5.3
Secondary CD8 lymphocyte rest	0.9	Astrocytes TNFalpha + IL-1beta	6.4
Secondary CD8 lymphocyte act	1.5	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	4.1	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	1.4	CCD1106 (Keratinocytes) none	4.8
LAK cells rest	15.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	11.0
LAK cells IL-2	6.9	Liver cirrhosis	3.8
LAK cells IL-2+IL-12	4.9	NCI-H292 none	19.3
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	11.2
LAK cells IL-2+ IL-18	2.5	NCI-H292 IL-9	20.7
LAK cells PMA/ionomycin	14.4	NCI-H292 IL-13	7.3
NK Cells IL-2 rest	2.2	NCI-H292 IFN gamma	5.7
Two Way MLR 3 day	1.5	HPAEC none	2.7
Two Way MLR 5 day	7.0	HPAEC TNF alpha + IL-1 beta	6.7
Two Way MLR 7 day	1.9	Lung fibroblast none	12.7
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	10.3

PBMC PWM	3.2	Lung fibroblast IL-4	11.1
PBMC PHA-L	7.6	Lung fibroblast IL-9	48.3
Ramos (B cell) none	5.8	Lung fibroblast IL-13	10.0
Ramos (B cell) ionomycin	7.4	Lung fibroblast IFN gamma	14.9
B lymphocytes PWM	3.2	Dermal fibroblast CCD1070 rest	20.6
B lymphocytes CD40L and IL-4	3.7	Dermal fibroblast CCD1070 TNF alpha	11.0
EOL-1 dbcAMP	2.5	Dermal fibroblast CCD1070 IL-1 beta	14.0
EOL-1 dbcAMP PMA/ionomycin	4.9	Dermal fibroblast IFN gamma	5.3
Dendritic cells none	12.2	Dermal fibroblast IL-4	7.4
Dendritic cells LPS	1.6	Dermal Fibroblasts rest	11.2
Dendritic cells anti-CD40	6.0	Neutrophils TNFa+LPS	6.2
Monocytes rest	1.2	Neutrophils rest	4.8
Monocytes LPS	**100.0**	Colon	0.0
Macrophages rest	7.9	Lung	0.0
Macrophages LPS	22.4	Thymus	0.0
HUVEC none	3.1	Kidney	5.1
HUVEC starved	5.7		

CNS_neurodegeneration_v1.0 Summary: Ag7623 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly down-regulated in the temporal cortex of Alzheimer's disease patients. Therefore, up-regulation of this gene or its protein product, or treatment with specific agonists for this receptor may be of use in reversing the dementia, memory loss, and neuronal death associated with this disease.

Panel 4.1D Summary: Ag7623 Highest expression of this gene is detected in LPS treated monocytes (CT=32.3). Expression of this gene is higher in the stimulted as compared to resting monocytes (CT=38). Thus, expression of this gene may be used to distinguish between activated and resting monocytes. In addition, upon activation with pathogens such as LPS, monocytes contribute to the innate and specific immunity by migrating to the site of tissue injury and releasing inflammatory cytokines. This release contributes to the inflammation process. Therefore, therapeutic modulation of the expression of this gene or the protein encoded by this gene may prevent the recruitment of monocytes and the initiation of the inflammatory process, and reduce the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, or rheumatoid arthritis.

In addition, low levels of expression of this gene are also seen in NCI-H292, coronery artery SMC, activated macrophage and lung fibroblasts. Therefore, therapeutic modulation of this gene or its protein product may be useful in the treatment of asthma, psoriasis, arthritis, allergy, chronic obstructive pulmonary disease, and emphysema.

J. CG144193-01: SECRETED PHOSPHOPROTEIN 24 PRECURSOR

Expression of full length physical clone CG144193-01 was assessed using the primer-probe set Ag7040, described in Table JA. Results of the RTQ-PCR runs are shown in Table JB.

Table JA. Probe Name Ag7040

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-actatgtgtccacgtctgagtctt-3'	24	326	260
Probe	TET-5'-atttatgagatcccaacatgtccccaaa-3'-TAMRA	28	370	261
Reverse	5'-tgagaccaaatagataattgcttctc-3'	26	399	262

Table JB. General_screening_panel_v1.6

Tissue Name	Rel. Exp.(%) Ag7040, Run 282273676	issue Name	Rel. Exp.(%) Ag7040, Run 282273676
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0

Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	20.6	Brain (Thalamus) Pool	0.0
Fetal Liver	**100.0**	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

General_screening_panel_v1.6 Summary: Ag7040 Significant expression is detected only in fetal liver (CT=33.8). Interestingly, this gene is expressed at much higher levels in fetal when compared to adult liver tissue (CT = 40). This observation suggests that expression of this gene can be used to differentiate between the fetal and adult sources of this tissue. In addition, the relative overexpression of this gene in fetal liver suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult to restore liver mass and/or function.

Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases, including cirrhosis and fibrosis.

K. CG144884-02: B-LYMPHOCYTE ACTIVATION MARKER BLAST-1 PRECURSOR

Expression of full-length physical clone CG144884-02 was assessed using the primer-probe set Ag4390, described in Table KA. Results of the RTQ-PCR runs are shown in Tables KB and KC.

Table KA. Probe Name Ag4390

Primers		Length	Start Position	SEQ ID No
Forward	5'-gtctggctctggaattgctact-3'	22	45	263
Probe	TET-5'-ctctgtcactcctggtgaccagcatt-3'-TAMRA	26	72	264
Reverse	5'-agaccacggtcatatgtaccaa-3'	22	107	265

Table KB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4390, Run 222641236	issue Name	Rel. Exp.(%) Ag4390, Run 222641236
Adipose	38.7	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	31.9
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.1
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	1.7	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	8.6	Colon ca. CaCo-2	0.0
Placenta	4.8	Colon cancer tissue	40.6
Uterus Pool	5.8	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	24.8
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	9.4
Ovarian ca. IGROV-1	0.0	Stomach Pool	13.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	19.6

Ovary	10.4	Fetal Heart	1.3
Breast ca. MCF-7	0.0	Heart Pool	5.6
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	21.9
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	1.8
Breast ca. T47D	0.0	Skeletal Muscle Pool	2.9
Breast ca. MDA-N	0.0	Spleen Pool	69.7
Breast Pool	14.2	Thymus Pool	**100.0**
Trachea	49.0	CNS cancer (glio/astro) U87-MG	0.4
Lung	0.9	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	31.9	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.1
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.4
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.3
Lung ca. NCI-H23	0.0	Brain (fetal)	0.8
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	1.4
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.9
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.8
Liver	5.7	Brain (Thalamus) Pool	1.0
Fetal Liver	14.0	Brain (whole)	1.5
Liver ca. HepG2	0.0	Spinal Cord Pool	6.4
Kidney Pool	17.7	Adrenal Gland	7.0
Fetal Kidney	3.2	Pituitary gland Pool	1.2
Renal ca. 786-0	0.0	Salivary Gland	14.7
Renal ca. A498	0.0	Thyroid (female)	14.1
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	21.2

Table KC. Panel 4.1D

Tissue Name	**Rel. Exp.() Ag4390, Run 186502193**	**Tissue Name**	**Rel. Exp.(%) Ag4390, Run 186502193**
Secondary Th1 act	40.3	HUVEC IL-1beta	0.0
Secondary Th2 act	33.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	22.7	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	19.1	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	18.8	HUVEC IL-11	2.0
Secondary Tr1 rest	24.0	Lung Microvascular EC none	0.2
Primary Th1 act	31.9	Lung Microvascular EC TNFalpha + IL-1beta	0.0

Primary Th2 act	18.2	Microvascular Dermal EC none	0.0
Primary Tr1 act	22.4	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	19.3	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	17.1	Small airway epithelium none	0.0
Primary Tr1 rest	31.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	22.1	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	41.2	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	32.5	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	20.4	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	24.3	KU-812 (Basophil) rest	0.7
CD4 lymphocyte none	22.4	KU-812 (Basophil) PMA/ionomycin	0.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	27.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	23.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	37.4	Liver cirrhosis	0.5
LAK cells IL-2+IL-12	38.2	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	35.1	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	30.4	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	22.8	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	44.8	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	28.3	HPAEC none	0.3
Two Way MLR 5 day	19.6	HPAEC TNF alpha + IL-1 beta	0.3
Two Way MLR 7 day	10.3	Lung fibroblast none	0.2
PBMC rest	23.3	Lung fibroblast TNF alpha + IL-1 beta	0.2
PBMC PWM	27.0	Lung fibroblast IL-4	0.1
PBMC PHA-L	12.8	Lung fibroblast IL-9	0.0
Ramos (B cell) none	16.5	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	31.4	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	21.2	Dermal fibroblast CCD1070 rest	1.3
B lymphocytes CD40L and IL-4	25.0	Dermal fibroblast CCD1070 TNF alpha	27.4
EOL-1 dbcAMP	1.1	Dermal fibroblast CCD1070 IL-1 beta	0.2
EOL-1 dbcAMP PMA/ionomycin	0.3	Dermal fibroblast IFN gamma	0.3
Dendritic cells none	14.6	Dermal fibroblast IL-4	0.7
Dendritic cells LPS	16.6	Dermal Fibroblasts rest	0.7
Dendritic cells anti-CD40	6.3	Neutrophils TNFa+LPS	18.9
Monocytes rest	39.5	Neutrophils rest	6.0

Monocytes LPS	**100.0**	Colon	2.3
Macrophages rest	12.2	Lung	1.7
Macrophages LPS	40.1	Thymus	9.5
HUVEC none	0.0	Kidney	1.2
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag4390 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag4390 Highest expression of this gene is detected in thymus (CT=29.4). The protein encoded for by this gene could therefore may play an important role in T cell development. Small molecule therapeutics, or antibody therapeutics designed against the protein encoded for by this gene could be utilized to modulate immune function (T cell development) and be important for organ transplant, AIDS treatment or post chemotherapy immune reconstitution.

Moderate to low levels of expression of this gene is also seen in tissues with metabolic/endocrine functions including pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Interestingly, this gene is expressed at much higher levels in fetal (CT=31) when compared to adult lung (CT=36). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung related diseases.

Panel 4.1D Summary: Ag4390 This gene appears to be expressed mainly in hematopoietic cells, including T cells, B cells, LAK cells, dendritic cells, monocytes and macrophages. This gene encodes a protein with homology to BLAST1, an activation-associated cell surface glycoprotein expressed primarily in mitogen-stimulated human lymphocytes. The expression of this gene in hematopoietic cells and thymus on Panel 1.4 is consistent with this characterization. Highest expression of this gene is seen in LPS treated monocytes (CT=26). Upon activation with pathogens such as LPS, monocytes contribute to the innate and specific immunity by migrating to the site of tissue injury and

releasing inflammatory cytokines. This release contributes to the inflammation process. Therefore, modulation of the expression of the protein encoded by this transcript may prevent the recruitment of monocytes and the initiation of the inflammatory process, and reduce the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, or rheumatoid arthritis.

L. CG145198-01: Novel Secreted Protein

Expression of full-length physical clone CG145198-01 was assessed using the primer-probe set Ag6943, described in Table LA. Results of the RTQ-PCR runs are shown in Table LB.

Table LA. Probe Name Ag6943

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-cccagaccagatgacctatctt-3'	22	299	266
Probe	TET-5'-ccttccagctctgagtcacttcccat-3'-TAMRA	26	321	267
Reverse	5'-aatggtctcagtgacttcgattaac-3'	25	358	268

Table LB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag6943, Run 278388849	issue Name	Rel. Exp.(%) Ag6943, Run 278388849
Adipose	9.9	Renal ca. TK-10	31.4
Melanoma* Hs688(A).T	19.9	Bladder	37.1
Melanoma* Hs688(B).T	17.4	Gastric ca. (liver met.) NCI-N87	36.9
Melanoma* M14	87.1	Gastric ca. KATO III	75.3
Melanoma* LOXIMVI	55.1	Colon ca. SW-948	7.4
Melanoma* SK-MEL-5	42.0	Colon ca. SW480	61.1
Squamous cell carcinoma SCC-4	21.6	Colon ca.* (SW480 met) SW620	23.3
Testis Pool	17.8	Colon ca. HT29	19.3
Prostate ca.* (bone met) PC-3	63.7	Colon ca. HCT-116	33.2
Prostate Pool	13.6	Colon ca. CaCo-2	29.1
Placenta	6.6	Colon cancer tissue	15.4
Uterus Pool	4.6	Colon ca. SW1116	7.7
Ovarian ca. OVCAR-3	18.6	Colon ca. Colo-205	4.0
Ovarian ca. SK-OV-3	32.1	Colon ca. SW-48	10.7

Ovarian ca. OVCAR-4	19.8	Colon Pool	17.6
Ovarian ca. OVCAR-5	27.9	Small Intestine Pool	12.4
Ovarian ca. IGROV-1	14.9	Stomach Pool	6.9
Ovarian ca. OVCAR-8	19.8	Bone Marrow Pool	9.6
Ovary	6.3	Fetal Heart	5.2
Breast ca. MCF-7	19.9	Heart Pool	4.4
Breast ca. MDA-MB-231	97.3	Lymph Node Pool	12.5
Breast ca. BT 549	**100.0**	Fetal Skeletal Muscle	3.5
Breast ca. T47D	9.6	Skeletal Muscle Pool	0.4
Breast ca. MDA-N	23.7	Spleen Pool	11.9
Breast Pool	11.4	Thymus Pool	73.7
Trachea	15.8	CNS cancer (glio/astro) U87-MG	22.7
Lung	2.1	CNS cancer (glio/astro) U-118-MG	63.3
Fetal Lung	16.2	CNS cancer (neuro;met) SK-N-AS	44.4
Lung ca. NCI-N417	8.0	CNS cancer (astro) SF-539	31.0
Lung ca. LX-1	13.1	CNS cancer (astro) SNB-75	59.5
Lung ca. NCI-H146	19.1	CNS cancer (glio) SNB-19	19.1
Lung ca. SHP-77	65.1	CNS cancer (glio) SF-295	59.5
Lung ca. A549	39.2	Brain (Amygdala) Pool	13.7
Lung ca. NCI-H526	5.4	Brain (cerebellum)	26.8
Lung ca. NCI-H23	33.2	Brain (fetal)	15.8
Lung ca. NCI-H460	11.3	Brain (Hippocampus) Pool	10.0
Lung ca. HOP-62	28.1	Cerebral Cortex Pool	14.2
Lung ca. NCI-H522	70.7	Brain (Substantia nigra) Pool	7.5
Liver	6.8	Brain (Thalamus) Pool	12.2
Fetal Liver	11.5	Brain (whole)	8.6
Liver ca. HepG2	21.3	Spinal Cord Pool	12.0
Kidney Pool	24.5	Adrenal Gland	12.9
Fetal Kidney	11.3	Pituitary gland Pool	3.2
Renal ca. 786-0	23.2	Salivary Gland	9.5
Renal ca. A498	8.5	Thyroid (female)	4.2
Renal ca. ACHN	11.0	Pancreatic ca. CAPAN2	15.1
Renal ca. UO-31	21.0	Pancreas Pool	11.1

General_screening_panel_v1.6 Summary: Ag6943 Highest expression of this gene is seen in a breast cancer cell line (CT=27.8). This gene is ubiquitously expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

In addition, this gene is expressed at much higher levels in fetal lung tissue (CT=30) when compared to expression in the adult counterpart (CT=33). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

M. CG145650-01 and CG145650-02: Lectin C-type Domain protein

Expression of full-length physical clones CG145650-01 and CG145650-02 was assessed using the primer-probe sets Ag6531, AG7094, Ag7397, and Ag7478, described in Tables MA, MB, MC, and MD. Results of the RTQ-PCR runs are shown in Tables ME, MF, and MG. Please note that Ag7094 is specific to CG145650-02 and Ag6531 and Ag7397 are specific to CG145650-01.

Table MA. Probe Name Ag6531

Primers	Sequencs	Length	Start Position	SEQ ID No
Forward	5'-agtagaaataaagtagcagttggaactaaa-3'	30	401	269
Probe	TET-5'-acttccaattctttgggcaacagctc-3'-TAMRA	26	433	270
Reverse	5'-cagcctcttctgcagagaca-3'	20	464	271

Table MB. Probe Name Ag7094

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-agacaccatacaatgatgttaattgtc-3'	27	636	272

Probe	TET-5'-tctcacaaactgacctttgaggacca-3'-TAMRA	26	664	273
Reverse	5'-agaatgttcagttcataagtggatctt-3'	27	695	274

Table MC. Probe Name Ag7397

Primers	Sequencs	Length	Start Position	SEQ ID No
Forward	5'-cttgccaagatgctgattca-3'	20	382	275
Probe	TET-5'-cagttggaactaaatgacttccaattcttt-3'-TAMRA	30	417	276
Reverse	5'-tctgcagagacagcctgga-3'	19	457	277

Table MD. Probe Name Ag7478

Primers	Sequencs	Length	Start Position	SEQ ID No
Forward	5'-ggaagtcatttagttccaactgcta-3'	25	414	278
Probe	TET-5'-atttctactgaatcagcatcttggcaagac-3'-TAMRA	30	380	279
Reverse	5'-aggtgagcctccattctagc-3'	20	345	280

Table MD. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag7094, Run 306266972	Rel. Exp.(%) Ag7397, Run 306266974	Rel. Exp.(%) g7478, Run 306518766	Tissue Name	Rel. Exp.(%) Ag7094, Run 306266972	Rel. Exp.(%) Ag7397, Run 306266974	Rel. Exp.(%) Ag7478, Run 306518766
110967 COPD-F	0.0	4.6	8.1	112427 Match Control Psoriasis-F	14.2	65.1	75.3
110980 COPD-F	2.3	3.9	11.3	112418 Psoriasis-M	0.0	7.6	13.5
110968 COPD-M	1.6	10.0	11.7	112723 Match Control Psoriasis-M	2.2	21.6	13.6
110977 COPD-M	6.0	47.0	39.2	112419 Psoriasis-M	0.0	10.6	20.9
110989 Emphysema-F	5.1	46.3	26.1	112424 Match Control Psoriasis-M	3.1	10.5	24.1

110992 Emphysema-F	0.0	6.8	6.3	112420 Psoriasis-M	2.0	35.8	43.2
110993 Emphysema-F	0.0	10.2	12.8	112425 Match Control Psoriasis-M	5.4	38.4	44.8
110994 Emphysema-F	0.0	3.7	6.8	104689 (MF) OA Bone-Backus	24.5	84.7	91.4
110995 Emphysema-F	0.0	25.5	19.9	104690 (MF) Adj "Normal" Bone-Backus	5.0	22.5	26.2
110996 Emphysema-F	0.0	3.0	2.4	104691 (MF) OA Synovium-Backus	11.6	37.1	39.8
110997 Asthma-M	0.0	2.2	7.7	104692 (BA) OA Cartilage-Backus	0.0	0.0	0.0
111001 Asthma-F	3.7	9.3	13.0	104694 (BA) OA Bone-Backus	28.3	**100.0**	**100.0**
111002 Asthma-F	2.5	14.6	25.2	104695 (BA) Adj "Normal" Bone-Backus	7.6	33.9	33.2
111003 Atopic Asthma-F	3.9	16.4	23.0	104696 (BA) OA Synovium-Backus	6.6	27.9	33.4
111004 Atopic Asthma-F	2.6	23.3	20.2	104700 (SS) OA Bone-Backus	15.6	38.2	35.4
111005 Atopic Asthma-F	3.8	13.1	12.5	104701 (SS) Adj "Normal" Bone-Backus	8.6	37.6	39.2
111006 Atopic Asthma-F	0.0	4.5	0.6	104702 (SS) OA Synovium-Backus	9.2	46.3	40.9

111417 Allergy-M	1.5	6.3	7.3	117093 OA Cartilage Rep7	2.3	9.5	13.6
112347 Allergy-M	0.0	0.0	0.5	112672 OA Bone5	1.6	11.3	23.8
112349 Normal Lung-F	0.0	0.0	0.1	112673 OA Synovium5	2.0	6.8	16.4
112357 Normal Lung-F	4.0	41.8	37.6	112674 OA Synovial Fluid cells5	0.0	15.1	19.3
112354 Normal Lung-M	5.4	17.9	21.2	117100 OA Cartilage Rep14	0.0	3.1	5.6
112374 Crohns-F	3.4	22.7	19.1	112756 OA Bone9	0.0	11.2	12.6
112389 Match Control Crohns-F	2.4	6.1	7.4	112757 OA Synovium9	0.0	7.0	9.8
112375 Crohns-F	2.0	13.5	13.8	112758 OA Synovial Fluid Cells9	0.0	7.2	16.0
112732 Match Control Crohns-F	6.2	31.9	32.5	117125 RA Cartilage Rep2	0.0	2.7	7.0
112725 Crohns-M	0.0	3.3	11.0	113492 Bone2 RA	2.7	4.9	15.6
112387 Match Control Crohns-M	0.0	3.6	5.8	113493 Synovium2 RA	0.0	2.2	6.6
112378 Crohns-M	0.0	0.1	0.9	113494 Syn Fluid Cells RA	0.0	4.3	13.4
112390 Match Control Crohns-M	3.7	27.7	25.2	113499 Cartilage4 RA	0.0	5.8	16.5
112726 Crohns-M	3.3	22.2	18.9	113500 Bone4 RA	1.6	4.5	19.3
112731 Match Control Crohns-M	4.8	18.8	23.3	113501 Synovium4 RA	3.7	3.7	17.9
112380 Ulcer Col-F	2.1	14.6	11.5	113502 Syn Fluid Cells4 RA	0.0	4.1	12.2

112734 Match Control Ulcer Col-F	12.9	59.5	73.2	113495 Cartilage3 RA	2.1	3.5	11.3
112384 Ulcer Col-F	3.4	16.6	20.9	113496 Bone3 RA	0.0	3.4	10.9
112737 Match Control Ulcer Col-F	0.0	6.0	7.9	113497 Synovium3 RA	1.7	2.3	6.7
112386 Ulcer Col-F	0.0	3.3	7.6	113498 Syn Fluid Cells3 RA	0.0	5.2	12.5
112738 Match Control Ulcer Col-F	0.0	8.2	7.7	117106 Normal Cartilage Rep20	0.0	0.4	1.2
112381 Ulcer Col-M	0.0	0.1	0.0	113663 Bone3 Normal	0.0	0.0	0.0
112735 Match Control Ulcer Col-M	0.0	1.2	3.1	113664 Synovium3 Normal	0.0	0.0	0.0
112382 Ulcer Col-M	0.0	6.0	11.3	113665 Syn Fluid Cells3 Normal	0.0	0.0	0.2
112394 Match Control Ulcer Col-M	0.0	1.9	4.4	117107 Normal Cartilage Rep22	0.0	1.3	4.1
112383 Ulcer Col-M	**100.0**	22.2	17.8	113667 Bone4 Normal	0.0	6.1	12.4
112736 Match Control Ulcer Col-M	0.0	5.6	6.2	113668 Synovium4 Normal	0.0	7.0	15.9
112423 Psoriasis-F	1.3	11.7	29.5	113669 Syn Fluid Cells4 Normal	0.0	16.5	15.8

Table ME. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag7397, Run 306066639	issue Name	Rel. Exp.(%) Ag7397, Run 306066639
Adipose	36.6	Renal ca. TK-10	9.2
Melanoma* Hs688(A).T	5.2	Bladder	91.4
Melanoma* Hs688(B).T	7.0	Gastric ca. (liver met.) NCI-N87	36.3
Melanoma* M14	0.0	Gastric ca. KATO III	12.6
Melanoma* LOXIMVI	2.8	Colon ca. SW-948	5.6
Melanoma* SK-MEL-5	11.4	Colon ca. SW480	30.6
Squamous cell carcinoma SCC-4	2.1	Colon ca.* (SW480 met) SW620	17.8
Testis Pool	29.3	Colon ca. HT29	2.3
Prostate ca.* (bone met) PC-3	11.3	Colon ca. HCT-116	20.9
Prostate Pool	4.4	Colon ca. CaCo-2	24.1
Placenta	0.0	Colon cancer tissue	100.0
Uterus Pool	12.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	3.8	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	8.5	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	42.0
Ovarian ca. OVCAR-5	48.3	Small Intestine Pool	69.3
Ovarian ca. IGROV-1	0.0	Stomach Pool	38.4
Ovarian ca. OVCAR-8	7.3	Bone Marrow Pool	12.2
Ovary	29.1	Fetal Heart	14.9
Breast ca. MCF-7	11.8	Heart Pool	8.4
Breast ca. MDA-MB-231	12.9	Lymph Node Pool	43.8
Breast ca. BT 549	24.3	Fetal Skeletal Muscle	15.7
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	70.7
Breast Pool	56.6	Thymus Pool	78.5
Trachea	15.4	CNS cancer (glio/astro) U87-MG	11.2
Lung	33.9	CNS cancer (glio/astro) U-118-MG	45.1
Fetal Lung	76.3	CNS cancer (neuro;met) SK-N-AS	14.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	9.1
Lung ca. LX-1	19.2	CNS cancer (astro) SNB-75	23.3
Lung ca. NCI-H146	4.3	CNS cancer (glio) SNB-19	4.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	44.4
Lung ca. A549	11.8	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	6.1	Brain (cerebellum)	11.1
Lung ca. NCI-H23	36.3	Brain (fetal)	28.1
Lung ca. NCI-H460	14.8	Brain (Hippocampus) Pool	21.0
Lung ca. HOP-62	12.7	Cerebral Cortex Pool	32.3
Lung ca. NCI-H522	5.6	Brain (Substantia nigra) Pool	22.4

Liver	0.0	Brain (Thalamus) Pool	27.5
Fetal Liver	6.2	Brain (whole)	14.1
Liver ca. HepG2	0.0	Spinal Cord Pool	28.9
Kidney Pool	0.0	Adrenal Gland	14.2
Fetal Kidney	2.0	Pituitary gland Pool	0.0
Renal ca. 786-0	11.3	Salivary Gland	0.0
Renal ca. A498	4.4	Thyroid (female)	10.4
Renal ca. ACHN	14.7	Pancreatic ca. CAPAN2	3.3
Renal ca. UO-31	5.2	Pancreas Pool	16.7

Table MF. Panel 4.1D

Tissue Name	Rel. Exp.(%) g7397, Run 305065214	Rel. Exp.(%) Ag7478, Run 306413263	Tissue Name	Rel. Exp.(%) Ag7397, Run 305065214	Rel. Exp.(%) Ag7478, Run 306413263
Secondary Th1 act	0.0	0.5	HUVEC IL-1beta	0.0	0.0
Secondary Th2 act	2.0	1.1	HUVEC IFN gamma	1.5	0.0
Secondary Tr1 act	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.4	0.0
Secondary Th2 rest	0.3	0.0	HUVEC IL-11	0.3	0.2
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	2.3	0.3
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0	0.0
Primary Th2 act	2.2	0.4	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	1.7	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	1.0	0.0
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0	0.0
Primary Th2 rest	0.0	0.0	Small airway epithelium none	0.0	0.0
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	1.9	0.0
CD45RA CD4 lymphocyte act	1.3	0.0	Coronery artery SMC rest	0.0	0.0
CD45RO CD4 lymphocyte act	0.7	7.4	Coronery artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.7	0.2	Astrocytes rest	0.3	0.0

Secondary CD8 lymphocyte rest	0.4	0.0	Astrocytes TNFalpha + IL-1beta	0.7	0.0
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.6	0.2	KU-812 (Basophil) PMA/ionomycin	2.1	1.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	1.5	0.0
LAK cells rest	37.1	8.6	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.4	0.0
LAK cells IL-2	0.0	0.0	Liver cirrhosis	1.9	1.4
LAK cells IL-2+IL-12	0.0	0.0	NCI-H292 none	1.5	1.9
LAK cells IL-2+IFN gamma	1.0	0.3	NCI-H292 IL-4	2.9	0.0
LAK cells IL-2+ IL-18	1.0	0.0	NCI-H292 IL-9	1.5	0.0
LAK cells PMA/ionomycin	40.6	73.7	NCI-H292 IL-13	0.3	2.6
NK Cells IL-2 rest	3.0	1.6	NCI-H292 IFN gamma	0.0	0.0
Two Way MLR 3 day	17.8	45.7	HPAEC none	0.0	0.0
Two Way MLR 5 day	3.3	0.0	HPAEC TNF alpha + IL-1 beta	0.0	3.5
Two Way MLR 7 day	0.0	0.0	Lung fibroblast none	3.0	2.4
PBMC rest	6.8	14.5	Lung fibroblast TNF alpha + IL-1 beta	2.9	0.0
PBMC PWM	0.0	7.0	Lung fibroblast IL-4	1.1	0.0
PBMC PHA-L	6.8	8.8	Lung fibroblast IL-9	2.0	4.4
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-13	0.2	0.0
Ramos (B cell) ionomycin	0.9	0.8	Lung fibroblast IFN gamma	2.5	0.6
B lymphocytes PWM	0.0	0.0	Dermal fibroblast CCD1070 rest	1.4	1.1
B lymphocytes CD40L and IL-4	23.8	35.6	Dermal fibroblast CCD1070 TNF alpha	0.4	1.4
EOL-1 dbcAMP	6.7	7.6	Dermal fibroblast CCD1070 IL-1 beta	0.5	0.0
EOL-1 dbcAMP PMA/ionomycin	3.8	0.5	Dermal fibroblast IFN gamma	1.2	0.0
Dendritic cells none	92.7	52.5	Dermal fibroblast IL-4	1.8	1.7
Dendritic cells LPS	31.2	20.7	Dermal Fibroblasts rest	1.0	0.0
Dendritic cells anti-CD40	62.0	75.8	Neutrophils TNFa+LPS	14.4	29.1
Monocytes rest	48.3	85.3	Neutrophils rest	**100.0**	**100.0**

Monocytes LPS	49.7	92.7	Colon	0.0	0.0
Macrophages rest	8.7	21.8	Lung	1.2	0.0
Macrophages LPS	6.6	23.0	Thymus	1.4	0.5
HUVEC none	0.0	0.0	Kidney	1.1	0.0
HUVEC starved	1.3	0.0			

AI_comprehensive panel_v1.0 Summary: Ag7397/Ag7478 Two experiments with two different probe and primer sets produce results that are in excellent agreement, with highest expression detected in an osteoarthritic bone sample (CTs=27-29). Low to moderate expression is seen in many of the samples on this panel, with slightly higher expression in clusters of samples derived from psoriasis and OA samples. Thus, this gene may be involved in the pathogenesis and/or treatment of these diseases.

Ag7094 Low levels of expression of this gene are detected in a single ulcerative colitis sample (CT=33.3). Interestingly, expression of this gene is higher in colitis sample as compared to the matched control sample (CT=40). Therefore, expression of this gene may be used as marker to detect the presence of ulcerative colitis and also, therapeutic modulation of this gene or its protein product may be useful in the treatment of ulcerative colitis.

General_screening_panel_v1.6 Summary: Ag7397 Detectable levels of expression are limited to samples from fetal lung, bladder, thymus, colon cancer, and small intestine (CTs=34-35). Ag6531 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4.1D Summary: Ag7397/Ag7478 Two experiments with two different probe and primer sets produce results that are in excellent agreement, with highest expression detected in resting neutrophils (CTs=30-31). In addition, prominent expression is seen in dendritic cells, macrophages, monocytes, and LAK cells. This transcript appears to be down-regulated in activated neutrophils (CTs=32-33), suggesting that the protein encoded by this gene is produced by resting neutrophils but not by activated neutrophils. Thus, expression of this gene could be used to differentiate between resting and activated neutrophils. Furthermore, the gene product may reduce activation of these inflammatory cells and be useful as a protein therapeutic to reduce or eliminate the symptoms in patients with Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis. In addition, modulation of this gene product may be effective in increasing the immune

response in patients with AIDS or other immunodeficiencies. Ag6531/Ag7094 Expression of this gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

N. CG145978-01: DUF221 domain containing membrane protein

Expression of gene CG145978-01 was assessed using the primer-probe set Ag7596, described in Table NA. Results of the RTQ-PCR runs are shown in Tables NB and NC.

Table NA. Probe Name Ag7596

Primers		Length	Start Position	SEQ ID No
Forward	5'-acagatgcagacagccatga-3'	20	250	281
Probe	TET-5'-tctcacctctgtctccagctccgttg-3'-TAMRA	26	282	282
Reverse	5'-cacattgtccctttggtcaaa-3'	21	310	283

Table NB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag7596, Run 311288611	issue Name	Rel. Exp.(%) Ag7596, Run 311288611
AD 1 Hippo	4.9	Control (Path) 3 Temporal Ctx	5.2
AD 2 Hippo	15.9	Control (Path) 4 Temporal Ctx	12.9
AD 3 Hippo	5.3	AD 1 Occipital Ctx	9.3
AD 4 Hippo	4.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	4.1
AD 6 Hippo	15.9	AD 4 Occipital Ctx	8.2
Control 2 Hippo	27.5	AD 5 Occipital Ctx	62.0
Control 4 Hippo	4.5	AD 6 Occipital Ctx	12.2
Control (Path) 3 Hippo	2.8	Control 1 Occipital Ctx	4.0
AD 1 Temporal Ctx	13.9	Control 2 Occipital Ctx	61.1
AD 2 Temporal Ctx	19.8	Control 3 Occipital Ctx	9.0
AD 3 Temporal Ctx	5.4	Control 4 Occipital Ctx	4.8
AD 4 Temporal Ctx	9.9	Control (Path) 1 Occipital Ctx	62.0
AD 5 Inf Temporal Ctx	36.3	Control (Path) 2 Occipital Ctx	6.3
AD 5 Sup Temporal Ctx	33.0	Control (Path) 3 Occipital Ctx	2.3
AD 6 Inf Temporal Ctx	29.1	Control (Path) 4 Occipital Ctx	11.3
AD 6 Sup Temporal Ctx	34.6	Control 1 Parietal Ctx	4.5
Control 1 Temporal Ctx	2.8	Control 2 Parietal Ctx	27.2
Control 2 Temporal Ctx	32.8	Control 3 Parietal Ctx	10.0
Control 3 Temporal Ctx	19.9	Control (Path) 1 Parietal Ctx	67.8
Control 3 Temporal Ctx	6.7	Control (Path) 2 Parietal Ctx	15.6

Control (Path) 1 Temporal Ctx	49.3	Control (Path) 3 Parietal Ctx	2.0
Control (Path) 2 Temporal Ctx	18.0	Control (Path) 4 Parietal Ctx	30.4

Table NC. Panel 4.1D

Tissue Name	**Rel. Exp.() Ag7596, Run 310113205**	**Tissue Name**	**Rel. Exp.(%) Ag7596, Run 310113205**
Secondary Th1 act	12.3	HUVEC IL-1beta	31.0
Secondary Th2 act	20.0	HUVEC IFN gamma	20.6
Secondary Tr1 act	9.3	HUVEC TNF alpha + IFN gamma	2.7
Secondary Th1 rest	3.4	HUVEC TNF alpha + IL4	10.0
Secondary Th2 rest	4.6	HUVEC IL-11	9.7
Secondary Tr1 rest	3.7	Lung Microvascular EC none	64.6
Primary Th1 act	3.3	Lung Microvascular EC TNFalpha + IL-1beta	25.7
Primary Th2 act	16.6	Microvascular Dermal EC none	6.5
Primary Tr1 act	16.8	Microsvasular Dermal EC TNFalpha + IL-1beta	14.7
Primary Th1 rest	2.0	Bronchial epithelium TNFalpha + IL1beta	46.3
Primary Th2 rest	0.6	Small airway epithelium none	42.6
Primary Tr1 rest	0.5	Small airway epithelium TNFalpha + IL-1beta	71.7
CD45RA CD4 lymphocyte act	27.7	Coronery artery SMC rest	16.0
CD45RO CD4 lymphocyte act	17.8	Coronery artery SMC TNFalpha + IL-1beta	30.4
CD8 lymphocyte act	9.3	Astrocytes rest	32.8
Secondary CD8 lymphocyte rest	7.6	Astrocytes TNFalpha + IL-1beta	69.3
Secondary CD8 lymphocyte act	5.5	KU-812 (Basophil) rest	59.9
CD4 lymphocyte none	2.2	KU-812 (Basophil) PMA/ionomycin	88.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	3.4	CCD1106 (Keratinocytes) none	72.7
LAK cells rest	4.6	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	35.1
LAK cells IL-2	6.8	Liver cirrhosis	9.9
LAK cells IL-2+IL-12	0.0	NCI-H292 none	42.9
LAK cells IL-2+IFN gamma	5.0	NCI-H292 IL-4	72.2
LAK cells IL-2+ IL-18	4.0	NCI-H292 IL-9	91.4
LAK cells PMA/ionomycin	36.3	NCI-H292 IL-13	50.7
NK Cells IL-2 rest	18.6	NCI-H292 IFN gamma	31.0
Two Way MLR 3 day	16.4	HPAEC none	17.3
Two Way MLR 5 day	5.7	HPAEC TNF alpha + IL-1 beta	64.6

Two Way MLR 7 day	4.4	Lung fibroblast none	36.6
PBMC rest	3.1	Lung fibroblast TNF alpha + IL-1 beta	29.9
PBMC PWM	7.9	Lung fibroblast IL-4	30.6
PBMC PHA-L	4.8	Lung fibroblast IL-9	33.0
Ramos (B cell) none	1.3	Lung fibroblast IL-13	15.2
Ramos (B cell) ionomycin	7.0	Lung fibroblast IFN gamma	27.0
B lymphocytes PWM	7.7	Dermal fibroblast CCD1070 rest	22.2
B lymphocytes CD40L and IL-4	17.1	Dermal fibroblast CCD1070 TNF alpha	46.0
EOL-1 dbcAMP	16.3	Dermal fibroblast CCD1070 IL-1 beta	19.2
EOL-1 dbcAMP PMA/ionomycin	18.0	Dermal fibroblast IFN gamma	16.8
Dendritic cells none	24.3	Dermal fibroblast IL-4	33.4
Dendritic cells LPS	**100.0**	Dermal Fibroblasts rest	21.3
Dendritic cells anti-CD40	12.1	Neutrophils TNFa+LPS	2.8
Monocytes rest	3.1	Neutrophils rest	11.8
Monocytes LPS	60.3	Colon	2.5
Macrophages rest	15.7	Lung	8.4
Macrophages LPS	28.1	Thymus	4.9
HUVEC none	12.9	Kidney	28.9
HUVEC starved	22.1		

CNS_neurodegeneration_v1.0 Summary: Ag7596 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain, including the cortex and hipppocampus. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Panel 4.1D Summary: Ag7956 Highest expression of this gene is seen in LPS treated dendritic cells (CT=31.8). Moderate levels of expression are seen in many samples on this panel and particularly in cells derived from the lung and skin including IL-4, IL-9, IL-13 and IFN gamma activated-NCI-H292 mucoepidermoid cells as well as untreated NCI-H292 cells, IL-4, IL-9, IL-13 and IFN gamma activated lung and dermal fibroblasts, human pulmonary aortic endothelial cells (treated and untreated), small airway epithelium (treated and untreated), treated bronchial epithelium and lung and dermal microvascular endothelial cells (treated and untreated). The expression of this gene in cells derived from or within the lung and skin suggests that this gene may be involved in normal conditions as

well as pathological and inflammatory lung and skin disorders that include chronic obstructive pulmonary disease, asthma, allergy, psoriasis and emphysema.

O. CG145997-01: Similar to Drosophila FRY gene

Expression of gene CG145997-01 was assessed using the primer-probe set Ag7557, described in Table OA.

Table OA. Probe Name Ag7557

Primers	Sequencs	Length	Start Position	SEQ ID No
Forward	5'-ctgagctcgagaaagaagcat-3'	21	976	284
Probe	TET-5'-cgagacattttcggatctttatttaatacc-3'-TAMRA	30	1002	285
Reverse	5'-atctatacaaagattccagtgcaact-3'	26	1032	286

CNS_neurodegeneration_v1.0 Summary: Ag7557 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag7557 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

P. CG146119-01: PAPILIN

Expression of gene CG146119-01 was assessed using the primer-probe set Ag7571, described in Table PA.

Table PA. Probe Name Ag7571

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-gcttctacagtaagtgtctggaacac-3'	26	2362	287
Probe	TET-5'-cactcactgggctcattctgctgg-3'-TAMRA	24	2403	288
Reverse	5'-gttgtcatagcaacagccaaac-3'	22	2439	289

CNS_neurodegeneration_v1.0 Summary: Ag7571 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag7571 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Q. CG146202-01: MEMBRANE-ASSOCIATED LECTIN TYPE-

Expression of full-length physical clone CG146202-01 was assessed using the primer-probe set Ag7047, described in Table QA. Results of the RTQ-PCR runs are shown in Table QB.

Table QA. Probe Name Ag7047

Primers		Length	Start Position	SEQ ID No
Forward	5'-tgcagtggaacgcctgt-3'	17	564	290
Probe	TET-5'-ctgtccctgggaatggacattcttcc-3'-TAMRA	26	588	291
Reverse	5'-gtgatggagtcgtgccagt-3'	19	650	292

Table QB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag7047, Run 282273803	issue Name	Rel. Exp.(%) Ag7047, Run 282273803
Adipose	68.3	Renal ca. TK-10	0.8
Melanoma* Hs688(A).T	0.0	Bladder	47.3
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.4	Colon ca. SW480	0.7
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	13.9	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	6.2	Colon ca. CaCo-2	0.0
Placenta	100.0	Colon cancer tissue	93.3
Uterus Pool	2.2	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	24.1
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	9.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	27.5
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	16.8
Ovary	20.7	Fetal Heart	4.8
Breast ca. MCF-7	0.0	Heart Pool	7.2
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	12.3

Breast ca. BT 549	0.0	Fetal Skeletal Muscle	24.7
Breast ca. T47D	0.0	Skeletal Muscle Pool	5.3
Breast ca. MDA-N	0.0	Spleen Pool	4.1
Breast Pool	14.7	Thymus Pool	35.1
Trachea	43.2	CNS cancer (glio/astro) U87-MG	0.0
Lung	3.3	CNS cancer (glio/astro) U-118-MG	0.4
Fetal Lung	51.4	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	1.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	20.6
Lung ca. NCI-H23	0.0	Brain (fetal)	5.5
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.4
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	3.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	2.3
Liver	10.3	Brain (Thalamus) Pool	1.3
Fetal Liver	55.5	Brain (whole)	5.9
Liver ca. HepG2	0.0	Spinal Cord Pool	2.7
Kidney Pool	32.1	Adrenal Gland	49.7
Fetal Kidney	9.0	Pituitary gland Pool	2.5
Renal ca. 786-0	0.0	Salivary Gland	19.5
Renal ca. A498	0.0	Thyroid (female)	6.3
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.6	Pancreas Pool	1.7

General_screening_panel_v1.6 Summary: Ag7047 Highest expression of this gene is detected in placenta (CT=29). Moderate to low levels of expression of this gene are also seen in tissues with metabolic/endocrine functions, including pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Moderate levels of expression are also seen in a sample derived from colon cancer. Thus, therapeutic modulation of the expression or function of this gene may be useful in the treatment of colon cancer.

In addition, moderate levels of expression of this gene are also detected in fetal brain and cerebellum. Thus, therapeutic modulation of this gene may be useful in the treatment of neurological disorders such as ataxia and autism.

Interestingly, this gene is expressed at much higher levels in fetal (CT=30) when compared to adult lung (CT=34). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung related diseases.

R. CG146250-02: novel membrane protein

Expression of full-length physical clone CG146250-02 was assessed using the primer-probe set Ag7566, described in Table RA. Results of the RTQ-PCR runs are shown in Table RB.

Table RA. Probe Name Ag7566

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-agcttccaccatcactttca-3'	20	198	293
Probe	TET-5'-cacatgccgtgtccaaggagctc-3'-TAMRA	23	218	294
Reverse	5'-gacaaagaggaagtcattatccagtag-3'	27	246	295

Table RB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7566, Run 308751128	issue Name	Rel. Exp.(%) Ag7566, Run 308751128
AD 1 Hippo	0.0	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	10.4	Control (Path) 4 Temporal Ctx	10.4
AD 3 Hippo	5.5	AD 1 Occipital Ctx	10.6
AD 4 Hippo	0.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	22.8	AD 3 Occipital Ctx	2.9
AD 6 Hippo	0.0	AD 4 Occipital Ctx	29.9
Control 2 Hippo	0.0	AD 5 Occipital Ctx	0.0
Control 4 Hippo	11.2	AD 6 Occipital Ctx	0.0
Control (Path) 3 Hippo	2.5	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	0.0	Control 2 Occipital Ctx	0.0
AD 2 Temporal Ctx	13.6	Control 3 Occipital Ctx	38.7
AD 3 Temporal Ctx	7.8	Control 4 Occipital Ctx	6.3

AD 4 Temporal Ctx	11.0	Control (Path) 1 Occipital Ctx	28.1
AD 5 Inf Temporal Ctx	29.9	Control (Path) 2 Occipital Ctx	12.8
AD 5 Sup Temporal Ctx	25.9	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	52.9	Control (Path) 4 Occipital Ctx	0.0
AD 6 Sup Temporal Ctx	100.0	Control 1 Parietal Ctx	0.0
Control 1 Temporal Ctx	11.2	Control 2 Parietal Ctx	61.6
Control 2 Temporal Ctx	12.4	Control 3 Parietal Ctx	11.7
Control 3 Temporal Ctx	31.4	Control (Path) 1 Parietal Ctx	48.0
Control 3 Temporal Ctx	12.4	Control (Path) 2 Parietal Ctx	0.0
Control (Path) 1 Temporal Ctx	14.9	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	17.3	Control (Path) 4 Parietal Ctx	12.7

CNS_neurodegeneration_v1.0 Summary: Ag7566 Low levels of expression of this gene is restricted to a sample derived from Alzheimer's patient (CT=34.5). Thus, expression of this gene may be useful in distinguishing this sample from other samples used in this panel.

Panel 4.1D Summary: Ag7566 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

S. CG146625-01: Type IIIa Membrane Protein

Expression of full-length physical clone CG146625-01 was assessed using the primer-probe set Ag7052, described in Table SA. Results of the RTQ-PCR runs are shown in Table SB.

Table SA. Probe Name Ag7052

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-tgagaacctgcagcatcaga-3'	20	279	296
Probe	TET-5'-atacggcagctgactgcaaacctcagc-3'-TAMRA	27	305	297
Reverse	5'-tcctggtggtgaaaggatgt-3'	20	360	298

Table SB. General_screening_panel_v1.6

Tissue Name	Rel. Exp.(%) Ag7052, Run 282273862	issue Name	Rel. Exp.(%) Ag7052, Run 282273862
Adipose	5.9	Renal ca. TK-10	55.5
Melanoma* Hs688(A).T	19.2	Bladder	13.5

Melanoma* Hs688(B).T	26.1	Gastric ca. (liver met.) NCI-N87	**100.0**
Melanoma* M14	37.1	Gastric ca. KATO III	40.3
Melanoma* LOXIMVI	13.7	Colon ca. SW-948	17.9
Melanoma* SK-MEL-5	23.8	Colon ca. SW480	47.3
Squamous cell carcinoma SCC-4	26.6	Colon ca.* (SW480 met) SW620	29.9
Testis Pool	37.6	Colon ca. HT29	38.7
Prostate ca.* (bone met) PC-3	51.4	Colon ca. HCT-116	55.1
Prostate Pool	9.9	Colon ca. CaCo-2	45.7
Placenta	17.4	Colon cancer tissue	14.4
Uterus Pool	4.5	Colon ca. SW1116	15.0
Ovarian ca. OVCAR-3	22.1	Colon ca. Colo-205	24.5
Ovarian ca. SK-OV-3	28.1	Colon ca. SW-48	24.5
Ovarian ca. OVCAR-4	14.3	Colon Pool	15.4
Ovarian ca. OVCAR-5	76.3	Small Intestine Pool	17.2
Ovarian ca. IGROV-1	29.3	Stomach Pool	7.7
Ovarian ca. OVCAR-8	47.3	Bone Marrow Pool	7.6
Ovary	12.0	Fetal Heart	4.7
Breast ca. MCF-7	46.3	Heart Pool	10.3
Breast ca. MDA-MB-231	64.2	Lymph Node Pool	24.1
Breast ca. BT 549	26.4	Fetal Skeletal Muscle	5.2
Breast ca. T47D	9.5	Skeletal Muscle Pool	3.2
Breast ca. MDA-N	9.4	Spleen Pool	9.0
Breast Pool	12.0	Thymus Pool	16.6
Trachea	18.6	CNS cancer (glio/astro) U87-MG	66.4
Lung	3.8	CNS cancer (glio/astro) U-118-MG	74.7
Fetal Lung	14.1	CNS cancer (neuro;met) SK-N-AS	29.1
Lung ca. NCI-N417	6.8	CNS cancer (astro) SF-539	23.8
Lung ca. LX-1	45.1	CNS cancer (astro) SNB-75	54.0
Lung ca. NCI-H146	7.0	CNS cancer (glio) SNB-19	30.6
Lung ca. SHP-77	26.4	CNS cancer (glio) SF-295	82.9
Lung ca. A549	35.8	Brain (Amygdala) Pool	10.1
Lung ca. NCI-H526	4.8	Brain (cerebellum)	42.3
Lung ca. NCI-H23	59.9	Brain (fetal)	11.7
Lung ca. NCI-H460	22.1	Brain (Hippocampus) Pool	11.7
Lung ca. HOP-62	44.4	Cerebral Cortex Pool	14.4
Lung ca. NCI-H522	45.7	Brain (Substantia nigra) Pool	10.1
Liver	17.6	Brain (Thalamus) Pool	12.8
Fetal Liver	35.4	Brain (whole)	13.9
Liver ca. HepG2	36.3	Spinal Cord Pool	11.3
Kidney Pool	27.4	Adrenal Gland	24.3
Fetal Kidney	11.4	Pituitary gland Pool	5.0
Renal ca. 786-0	48.6	Salivary Gland	13.1
Renal ca. A498	7.0	Thyroid (female)	19.3
Renal ca. ACHN	25.9	Pancreatic ca. CAPAN2	62.4

Renal ca. UO-31	39.8	Pancreas Pool	9.7

General_screening_panel_v1.6 Summary: Ag7052 Highest expression of this gene is seen in a gastric cancer cell line (CT=28). This gene is widely expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

T. CG146625-02: Type IIIa Membrane Protein

Expression of full-length physical clone CG146625-02 was assessed using the primer-probe set Ag6939, described in Table TA. Results of the RTQ-PCR runs are shown in Table TB.

Table TA. Probe Name Ag6939

Primers		Length	Start Position	SEQ ID No
Forward	5'-gctgagccttccacgagtt-3'	19	680	299
Probe	TET-5'-tcatcccagatcaccatgcagaagc-3'-TAMRA	25	740	300
Reverse	5'-gtgctgaggtttgcagtcag-3'	20	809	301

Table TB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag6939, Run 278700426	issue Name	Rel. Exp.(%) Ag6939, Run 278700426
Adipose	8.7	Renal ca. TK-10	59.9
Melanoma* Hs688(A).T	47.6	Bladder	13.4
Melanoma* Hs688(B).T	36.9	Gastric ca. (liver met.) NCI-N87	88.3
Melanoma* M14	38.7	Gastric ca. KATO III	39.8
Melanoma* LOXIMVI	18.2	Colon ca. SW-948	5.6
Melanoma* SK-MEL-5	22.2	Colon ca. SW480	55.5
Squamous cell carcinoma SCC-4	54.3	Colon ca.* (SW480 met) SW620	25.7
Testis Pool	24.3	Colon ca. HT29	36.3
Prostate ca.* (bone met) PC-3	60.3	Colon ca. HCT-116	35.4
Prostate Pool	15.7	Colon ca. CaCo-2	32.8
Placenta	12.4	Colon cancer tissue	22.5
Uterus Pool	7.9	Colon ca. SW1116	19.3
Ovarian ca. OVCAR-3	29.9	Colon ca. Colo-205	21.8
Ovarian ca. SK-OV-3	27.2	Colon ca. SW-48	28.7
Ovarian ca. OVCAR-4	14.8	Colon Pool	20.6
Ovarian ca. OVCAR-5	50.7	Small Intestine Pool	16.5
Ovarian ca. IGROV-1	37.9	Stomach Pool	10.8
Ovarian ca. OVCAR-8	52.1	Bone Marrow Pool	5.3
Ovary	12.6	Fetal Heart	2.4
Breast ca. MCF-7	47.3	Heart Pool	11.9
Breast ca. MDA-MB-231	100.0	Lymph Node Pool	53.2
Breast ca. BT 549	23.0	Fetal Skeletal Muscle	1.1
Breast ca. T47D	12.1	Skeletal Muscle Pool	1.5
Breast ca. MDA-N	16.3	Spleen Pool	8.1
Breast Pool	25.3	Thymus Pool	11.8
Trachea	18.0	CNS cancer (glio/astro) U87-MG	54.7
Lung	8.2	CNS cancer (glio/astro) U-118-MG	70.7
Fetal Lung	12.2	CNS cancer (neuro;met) SK-N-AS	20.7
Lung ca. NCI-N417	8.1	CNS cancer (astro) SF-539	25.2
Lung ca. LX-1	22.1	CNS cancer (astro) SNB-75	39.8
Lung ca. NCI-H146	7.1	CNS cancer (glio) SNB-19	33.9
Lung ca. SHP-77	18.7	CNS cancer (glio) SF-295	77.4
Lung ca. A549	29.9	Brain (Amygdala) Pool	7.6
Lung ca. NCI-H526	5.0	Brain (cerebellum)	17.3
Lung ca. NCI-H23	8.2	Brain (fetal)	8.3
Lung ca. NCI-H460	23.0	Brain (Hippocampus) Pool	12.4
Lung ca. HOP-62	39.8	Cerebral Cortex Pool	10.0
Lung ca. NCI-H522	36.3	Brain (Substantia nigra) Pool	11.7

Liver	11.6	Brain (Thalamus) Pool	11.7
Fetal Liver	41.2	Brain (whole)	10.4
Liver ca. HepG2	23.7	Spinal Cord Pool	9.9
Kidney Pool	39.2	Adrenal Gland	24.0
Fetal Kidney	6.1	Pituitary gland Pool	6.4
Renal ca. 786-0	39.0	Salivary Gland	12.2
Renal ca. A498	4.6	Thyroid (female)	23.0
Renal ca. ACHN	20.6	Pancreatic ca. CAPAN2	52.5
Renal ca. UO-31	33.2	Pancreas Pool	10.3

General_screening_panel_v1.6 Summary: Ag6939 Highest expression of this gene is detected in a breast cancer MDA-MB-231 cell line (CT=32). Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Low levels of expression of this gene is also seen in samples derived from normal tissues represented by testis, prostate, ovary, trachea, fetal liver, colon, small intestine, lymph node, cerebellum, thyroid and adrenal gland. Therefore, therapeutic modulation of this gene or its protein product may be useful in the treatment of diseases related to these tissues.

U. CG147284-01: CADHERIN-6 PRECURSOR

Expression of full-length physical clone CG147284-01 was assessed using the primer-probe set Ag7567, described in Table UA.

Table UA. Probe Name Ag7567

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cgtgttgtctttgttgtcttga-3'	22	285	302
Probe	TET-5'-tgtgggcaagttacattcaaactttacca-3'-TAMRA	29	255	303
Reverse	5'-gaatacacaggatccgattatcagta-3'	26	229	304

CNS_neurodegeneration_v1.0 Summary: Ag7567 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag7567 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

V. CG148221-01 and CG148221-02: claudin domain containing novel TmMP

Expression of gene CG148221-01 and full-length physical clone was assessed using the primer-probe set Ag5625, described in Table VA. Results of the RTQ-PCR runs are shown in Tables VB, VC and VD.

Table VA. Probe Name Ag5625

Primers		Length	Start Position	SEQ ID No
Forward	5'-tttctgctggcagacatgat-3'	20	469	305
Probe	TET-5'-agcaccgacgccatcagtggatt-3'-TAMRA	23	496	306
Reverse	5'-caggctgcagtcacagaca-3'	19	526	307

Table VB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag5625, Run 244647005	issue Name	Rel. Exp.(%) Ag5625, Run 244647005
AD 1 Hippo	23.5	Control (Path) 3 Temporal Ctx	20.4
AD 2 Hippo	62.4	Control (Path) 4 Temporal Ctx	44.8
AD 3 Hippo	2.8	AD 1 Occipital Ctx	11.7
AD 4 Hippo	10.5	AD 2 Occipital Ctx (Missing)	3.0
AD 5 hippo	45.7	AD 3 Occipital Ctx	4.7
AD 6 Hippo	76.8	AD 4 Occipital Ctx	20.7
Control 2 Hippo	30.4	AD 5 Occipital Ctx	37.9
Control 4 Hippo	10.4	AD 6 Occipital Ctx	40.6
Control (Path) 3 Hippo	17.9	Control 1 Occipital Ctx	8.5
AD 1 Temporal Ctx	27.2	Control 2 Occipital Ctx	52.1
AD 2 Temporal Ctx	57.8	Control 3 Occipital Ctx	17.7
AD 3 Temporal Ctx	7.9	Control 4 Occipital Ctx	15.6
AD 4 Temporal Ctx	24.5	Control (Path) 1 Occipital Ctx	52.1
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	5.3
AD 5 SupTemporal Ctx	69.7	Control (Path) 3 Occipital Ctx	10.5

AD 6 Inf Temporal Ctx	74.7	Control (Path) 4 Occipital Ctx	20.4
AD 6 Sup Temporal Ctx	57.8	Control 1 Parietal Ctx	14.7
Control 1 Temporal Ctx	24.1	Control 2 Parietal Ctx	54.0
Control 2 Temporal Ctx	33.9	Control 3 Parietal Ctx	31.6
Control 3 Temporal Ctx	19.8	Control (Path) 1 Parietal Ctx	54.0
Control 4 Temporal Ctx	18.2	Control (Path) 2 Parietal Ctx	15.7
Control (Path) 1 Temporal Ctx	34.9	Control (Path) 3 Parietal Ctx	15.6
Control (Path) 2 Temporal Ctx	44.1	Control (Path) 4 Parietal Ctx	40.1

Table VC. General_screening_panel_v1.5

Tissue Name	**Rel. Exp.(%) Ag5625, Run 244646965**	**issue Name**	**Rel. Exp.(%) Ag5625, Run 244646965**
Adipose	18.4	Renal ca. TK-10	29.5
Melanoma* Hs688(A).T	3.7	Bladder	22.2
Melanoma* Hs688(B).T	2.4	Gastric ca. (liver met.) NCI-N87	12.2
Melanoma* M14	14.5	Gastric ca. KATO III	23.0
Melanoma* LOXIMVI	1.7	Colon ca. SW-948	12.0
Melanoma* SK-MEL-5	4.8	Colon ca. SW480	82.9
Squamous cell carcinoma SCC-4	4.0	Colon ca.* (SW480 met) SW620	36.6
Testis Pool	30.6	Colon ca. HT29	24.0
Prostate ca.* (bone met) PC-3	7.8	Colon ca. HCT-116	31.2
Prostate Pool	2.9	Colon ca. CaCo-2	31.2
Placenta	0.4	Colon cancer tissue	8.1
Uterus Pool	2.6	Colon ca. SW1116	4.2
Ovarian ca. OVCAR-3	2.8	Colon ca. Colo-205	2.8
Ovarian ca. SK-OV-3	8.2	Colon ca. SW-48	4.6
Ovarian ca. OVCAR-4	10.6	Colon Pool	5.6
Ovarian ca. OVCAR-5	23.2	Small Intestine Pool	12.3
Ovarian ca. IGROV-1	13.7	Stomach Pool	7.8
Ovarian ca. OVCAR-8	6.0	Bone Marrow Pool	6.7
Ovary	3.3	Fetal Heart	2.1
Breast ca. MCF-7	8.1	Heart Pool	3.0
Breast ca. MDA-MB-231	29.3	Lymph Node Pool	9.3
Breast ca. BT 549	7.1	Fetal Skeletal Muscle	4.1
Breast ca. T47D	4.3	Skeletal Muscle Pool	4.9
Breast ca. MDA-N	22.1	Spleen Pool	5.0
Breast Pool	6.0	Thymus Pool	9.3
Trachea	1.8	CNS cancer (glio/astro) U87-MG	30.8
Lung	16.6	CNS cancer (glio/astro) U-118-MG	28.1
Fetal Lung	3.7	CNS cancer (neuro;met) SK-N-AS	11.9

Lung ca. NCI-N417	6.4	CNS cancer (astro) SF-539	3.0
Lung ca. LX-1	**100.0**	CNS cancer (astro) SNB-75	25.9
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	17.1
Lung ca. SHP-77	18.4	CNS cancer (glio) SF-295	24.0
Lung ca. A549	51.4	Brain (Amygdala) Pool	8.7
Lung ca. NCI-H526	7.0	Brain (cerebellum)	28.5
Lung ca. NCI-H23	15.3	Brain (fetal)	11.3
Lung ca. NCI-H460	15.7	Brain (Hippocampus) Pool	10.2
Lung ca. HOP-62	3.3	Cerebral Cortex Pool	11.4
Lung ca. NCI-H522	29.5	Brain (Substantia nigra) Pool	11.6
Liver	1.7	Brain (Thalamus) Pool	14.3
Fetal Liver	8.5	Brain (whole)	7.3
Liver ca. HepG2	57.8	Spinal Cord Pool	7.9
Kidney Pool	24.1	Adrenal Gland	6.7
Fetal Kidney	7.4	Pituitary gland Pool	3.0
Renal ca. 786-0	10.4	Salivary Gland	6.8
Renal ca. A498	18.9	Thyroid (female)	5.2
Renal ca. ACHN	1.7	Pancreatic ca. CAPAN2	34.9
Renal ca. UO-31	2.5	Pancreas Pool	6.4

Table VD. Panel 4.1D

Tissue Name	**Rel. Exp.() Ag5625, Run 246490692**	**Tissue Name**	**Rel. Exp.(%) Ag5625, Run 246490692**
Secondary Th1 act	17.1	HUVEC IL-1beta	1.8
Secondary Th2 act	30.4	HUVEC IFN gamma	0.0
Secondary Tr1 act	15.2	HUVEC TNF alpha + IFN gamma	0.7
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	2.6
Secondary Tr1 rest	1.3	Lung Microvascular EC none	1.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	15.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	10.2	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	1.7
Primary Th2 rest	0.0	Small airway epithelium none	1.8
Primary Tr1 rest	0.5	Small airway epithelium TNFalpha + IL-1beta	1.1
CD45RA CD4 lymphocyte act	5.6	Coronery artery SMC rest	0.0

CD45RO CD4 lymphocyte act	9.7	Coronery artery SMC TNFalpha + IL-1beta	1.6
CD8 lymphocyte act	1.3	Astrocytes rest	0.6
Secondary CD8 lymphocyte rest	4.8	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	6.7	KU-812 (Basophil) rest	1.3
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	2.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.1	CCD1106 (Keratinocytes) none	2.0
LAK cells rest	3.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.0
LAK cells IL-2	8.5	Liver cirrhosis	2.6
LAK cells IL-2+IL-12	2.6	NCI-H292 none	1.7
LAK cells IL-2+IFN gamma	5.8	NCI-H292 IL-4	0.7
LAK cells IL-2+ IL-18	2.0	NCI-H292 IL-9	6.0
LAK cells PMA/ionomycin	16.4	NCI-H292 IL-13	1.8
NK Cells IL-2 rest	100.0	NCI-H292 IFN gamma	0.6
Two Way MLR 3 day	3.6	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.8
Two Way MLR 7 day	1.6	Lung fibroblast none	1.3
PBMC rest	0.7	Lung fibroblast TNF alpha + IL-1 beta	1.3
PBMC PWM	2.0	Lung fibroblast IL-4	1.4
PBMC PHA-L	1.3	Lung fibroblast IL-9	0.7
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	4.2	Lung fibroblast IFN gamma	0.6
B lymphocytes PWM	7.4	Dermal fibroblast CCD1070 rest	1.7
B lymphocytes CD40L and IL-4	6.0	Dermal fibroblast CCD1070 TNF alpha	31.4
EOL-1 dbcAMP	14.2	Dermal fibroblast CCD1070 IL-1 beta	0.3
EOL-1 dbcAMP PMA/ionomycin	1.1	Dermal fibroblast IFN gamma	1.0
Dendritic cells none	3.2	Dermal fibroblast IL-4	1.9
Dendritic cells LPS	1.3	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.6
Monocytes rest	0.5	Neutrophils rest	2.7
Monocytes LPS	3.5	Colon	0.7
Macrophages rest	0.7	Lung	0.0
Macrophages LPS	0.7	Thymus	0.0
HUVEC none	1.2	Kidney	2.3
HUVEC starved	2.5		

CNS_neurodegeneration_v1.0 Summary: Ag5625 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms

the expression of this gene at moderate levels in the brain. Please see Panel 1.5 for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.5 Summary: Ag5625 Highest expression of this gene is seen in a lung cancer cell line (CT=29.4). This gene is widely expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. This gene encodes a protein with homology to claudin, a family of proteins that are integral components of the tight junction. Members of this family have been shown to be upregulated in pancreatic cancer and colon cancer and in the former case proposed as novel targets for the treatment of this disease (Michl P. Gastroenterology 2001 Sep;121(3):678-84; Miwa, N. Oncol Res 2001;12(11-12):469-76) Therefore, therapeutic modulation of the expression or function of this protein may be of use in the treatment of these cancers.

Among tissues with metabolic function, this gene is expressed at low but significant levels in pituitary, adipose, adrenal gland, pancreas, thyroid, fetal liver and adult and fetal skeletal muscle and heart. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that disregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at low but significant levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Claudin 11 has been shown to be a component of the CNS myelin and has been implicated in the regulation of growth and differentiation via signal transduction pathways. Furthermore, evidence has been presented that shows that claudin 11 may be involved in the autoantigen that is responsible for the development of autoimmune demyelinating disease.(Bronstein JM. J Neurosci Res 2000 Mar 15;59(6):706-11). Therefore, therapeutic modulation of the expression or function of this putative claudin may be of use in the treatment of demyelinating diseases such as multiple sclerosis and in restoring normal function to the CNS.

Panel 4.1D Summary: Ag5625 Highest expression of this gene is seen in IL-2 treated NK cells (CT=29). This observation suggests that therapeutic modulation of this gene could be of use in the treatment of viral or bacterial intracellular infections.

W. CG149332-01: INTERFERON INDUCED TRANSMEMBRANE PROTEIN 3 (1-8U) - Like

Expression of gene CG149332-01 was assessed using the primer-probe set Ag7580, described in Table WA. Results of the RTQ-PCR runs are shown in Tables WB and WC.

Table WA. Probe Name Ag7580

Primers		Length	Start Position	SEQ ID No
Forward	5'-gagaagcatgaggtggctgt-3'	20	76	308
Probe	TET-5'-accccacaaccctgtgcctccag-3'-TAMRA	23	105	309
Reverse	5'-gcagatgtggatcatggtga-3'	20	131	310

Table WB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7580, Run 308752173	issue Name	Rel. Exp.(%) Ag7580, Run 308752173
AD 1 Hippo	17.2	Control (Path) 3 Temporal Ctx	17.4
AD 2 Hippo	30.4	Control (Path) 4 Temporal Ctx	27.7
AD 3 Hippo	3.7	AD 1 Occipital Ctx	12.9
AD 4 Hippo	29.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	24.3	AD 3 Occipital Ctx	0.0
AD 6 Hippo	**100.0**	AD 4 Occipital Ctx	14.8
Control 2 Hippo	8.7	AD 5 Occipital Ctx	7.0
Control 4 Hippo	40.1	AD 6 Occipital Ctx	3.5
Control (Path) 3 Hippo	7.5	Control 1 Occipital Ctx	8.2
AD 1 Temporal Ctx	7.9	Control 2 Occipital Ctx	12.1
AD 2 Temporal Ctx	23.2	Control 3 Occipital Ctx	13.8
AD 3 Temporal Ctx	5.7	Control 4 Occipital Ctx	7.5
AD 4 Temporal Ctx	11.8	Control (Path) 1 Occipital Ctx	16.4
AD 5 Inf Temporal Ctx	44.8	Control (Path) 2 Occipital Ctx	3.5
AD 5 Sup Temporal Ctx	87.7	Control (Path) 3 Occipital Ctx	3.2
AD 6 Inf Temporal Ctx	56.6	Control (Path) 4 Occipital Ctx	10.2
AD 6 Sup Temporal Ctx	56.3	Control 1 Parietal Ctx	4.1
Control 1 Temporal Ctx	14.4	Control 2 Parietal Ctx	62.0
Control 2 Temporal Ctx	28.7	Control 3 Parietal Ctx	2.9
Control 3 Temporal Ctx	28.5	Control (Path) 1 Parietal Ctx	38.7
Control 3 Temporal Ctx	6.1	Control (Path) 2 Parietal Ctx	13.4
Control (Path) 1 Temporal Ctx	40.3	Control (Path) 3 Parietal Ctx	3.7
Control (Path) 2 Temporal Ctx	53.2	Control (Path) 4 Parietal Ctx	29.1

Table WC. Panel 4.1D

Tissue Name	Rel. Exp.() Ag7580, Run 308748674	Tissue Name	Rel. Exp.(%) Ag7580, Run 308748674
Secondary Th1 act	9.5	HUVEC IL-1beta	2.5
Secondary Th2 act	21.3	HUVEC IFN gamma	19.8
Secondary Tr1 act	4.2	HUVEC TNF alpha + IFN gamma	0.7
Secondary Th1 rest	3.0	HUVEC TNF alpha + IL4	1.4
Secondary Th2 rest	11.1	HUVEC IL-11	2.9
Secondary Tr1 rest	15.1	Lung Microvascular EC none	2.4
Primary Th1 act	1.8	Lung Microvascular EC TNFalpha + IL-1beta	1.1
Primary Th2 act	25.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	12.1	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	12.4	Bronchial epithelium TNFalpha + IL1beta	10.1
Primary Th2 rest	21.2	Small airway epithelium none	2.0
Primary Tr1 rest	2.8	Small airway epithelium TNFalpha + IL-1beta	14.5
CD45RA CD4 lymphocyte act	21.0	Coronery artery SMC rest	5.5
CD45RO CD4 lymphocyte act	63.3	Coronery artery SMC TNFalpha + IL-1beta	7.9
CD8 lymphocyte act	15.1	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	10.7	Astrocytes TNFalpha + IL-1beta	2.1
Secondary CD8 lymphocyte act	1.5	KU-812 (Basophil) rest	3.3
CD4 lymphocyte none	19.3	KU-812 (Basophil) PMA/ionomycin	8.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	9.9	CCD1106 (Keratinocytes) none	13.7
LAK cells rest	33.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	8.6
LAK cells IL-2	12.3	Liver cirrhosis	3.5
LAK cells IL-2+IL-12	2.1	NCI-H292 none	19.3
LAK cells IL-2+IFN gamma	20.6	NCI-H292 IL-4	20.7
LAK cells IL-2+ IL-18	11.1	NCI-H292 IL-9	31.6
LAK cells PMA/ionomycin	14.5	NCI-H292 IL-13	22.2
NK Cells IL-2 rest	**100.0**	NCI-H292 IFN gamma	5.3
Two Way MLR 3 day	31.9	HPAEC none	1.2
Two Way MLR 5 day	4.9	HPAEC TNF alpha + IL-1 beta	5.9
Two Way MLR 7 day	5.7	Lung fibroblast none	1.1

PBMC rest	6.1	Lung fibroblast TNF alpha + IL-1 beta	1.1
PBMC PWM	9.9	Lung fibroblast IL-4	0.6
PBMC PHA-L	11.2	Lung fibroblast IL-9	2.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.9
Ramos (B cell) ionomycin	1.9	Lung fibroblast IFN gamma	4.2
B lymphocytes PWM	5.9	Dermal fibroblast CCD1070 rest	16.2
B lymphocytes CD40L and IL-4	60.3	Dermal fibroblast CCD1070 TNF alpha	59.9
EOL-1 dbcAMP	2.8	Dermal fibroblast CCD1070 IL-1 beta	10.6
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	1.5
Dendritic cells none	4.0	Dermal fibroblast IL-4	2.3
Dendritic cells LPS	1.8	Dermal Fibroblasts rest	0.4
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	6.3
Monocytes rest	0.3	Neutrophils rest	5.5
Monocytes LPS	4.7	Colon	0.6
Macrophages rest	2.0	Lung	0.6
Macrophages LPS	0.6	Thymus	1.2
HUVEC none	2.5	Kidney	4.2
HUVEC starved	1.4		

CNS_neurodegeneration_v1.0 Summary: Ag7580 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain, including the hippocampus and cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Panel 4.1D Summary: Ag7580 Highest expression of this gene is seen in IL-2 treated NK cells. Moderate to low levels of expression are seen in many samples on this panel, inlucding TNF-a treated and resting dermal fibroblasts, TNF-a and LPS treated neutrophils, activated primary and secondary T cells, and LAK cells. This expression suggests that modulation of the expression or function of this gene may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

X. CG149649-01: Type IIIA membrane protein

Expression of gene CG149649-01 was assessed using the primer-probe set Ag7568, described in Table XA. Results of the RTQ-PCR runs are shown in Tables XB and XC.

Table XA. Probe Name Ag7568

Primers	Sequenes	Length	Start Position	SEQ ID No
Forward	5'-ggcctcttggccctctact-3'	19	226	311
Probe	TET-5'-cctcctgcttttgctcttctggatctacag-3'-TAMRA	30	246	312
Reverse	5'-tatagcacccctgtgggagt-3'	20	290	313

Table XB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7568, Run 308751131	issue Name	Rel. Exp.(%) Ag7568, Run 308751131
AD 1 Hippo	18.7	Control (Path) 3 Temporal Ctx	10.4
AD 2 Hippo	40.6	Control (Path) 4 Temporal Ctx	20.2
AD 3 Hippo	12.0	AD 1 Occipital Ctx	15.1
AD 4 Hippo	8.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	60.7	AD 3 Occipital Ctx	11.0
AD 6 Hippo	100.0	AD 4 Occipital Ctx	26.8
Control 2 Hippo	41.2	AD 5 Occipital Ctx	32.1
Control 4 Hippo	29.7	AD 6 Occipital Ctx	19.5
Control (Path) 3 Hippo	14.2	Control 1 Occipital Ctx	9.9
AD 1 Temporal Ctx	20.2	Control 2 Occipital Ctx	55.5
AD 2 Temporal Ctx	50.7	Control 3 Occipital Ctx	16.6
AD 3 Temporal Ctx	10.7	Control 4 Occipital Ctx	12.0
AD 4 Temporal Ctx	28.7	Control (Path) 1 Occipital Ctx	68.8
AD 5 Inf Temporal Ctx	60.3	Control (Path) 2 Occipital Ctx	17.8
AD 5 Sup Temporal Ctx	47.3	Control (Path) 3 Occipital Ctx	6.1
AD 6 Inf Temporal Ctx	76.3	Control (Path) 4 Occipital Ctx	11.9
AD 6 Sup Temporal Ctx	82.9	Control 1 Parietal Ctx	16.5
Control 1 Temporal Ctx	10.7	Control 2 Parietal Ctx	43.8
Control 2 Temporal Ctx	42.6	Control 3 Parietal Ctx	15.1
Control 3 Temporal Ctx	21.5	Control (Path) 1 Parietal Ctx	59.0
Control 3 Temporal Ctx	13.0	Control (Path) 2 Parietal Ctx	21.2
Control (Path) 1 Temporal Ctx	39.2	Control (Path) 3 Parietal Ctx	10.2
Control (Path) 2 Temporal Ctx	37.4	Control (Path) 4 Parietal Ctx	21.5

Table XC. Panel 4.1D

Tissue Name	Rel. Ep.(%) Ag7568, Run 308748452	Tissue Name	Rel. Exp.(%) Ag7568, Run 308748452
Secondary Th1 act	45.4	HUVEC IL-1beta	39.5
Secondary Th2 act	70.2	HUVEC IFN gamma	47.3
Secondary Tr1 act	26.8	HUVEC TNF alpha + IFN gamma	13.6
Secondary Th1 rest	4.5	HUVEC TNF alpha + IL4	10.5
Secondary Th2 rest	5.5	HUVEC IL-11	19.1
Secondary Tr1 rest	10.4	Lung Microvascular EC none	51.1
Primary Th1 act	5.4	Lung Microvascular EC TNFalpha + IL-1beta	22.1
Primary Th2 act	38.2	Microvascular Dermal EC none	6.0
Primary Tr1 act	33.9	Microsvasular Dermal EC TNFalpha + IL-1beta	14.7
Primary Th1 rest	2.4	Bronchial epithelium TNFalpha + IL1beta	6.8
Primary Th2 rest	4.5	Small airway epithelium none	27.0
Primary Tr1 rest	0.8	Small airway epithelium TNFalpha + IL-1beta	39.8
CD45RA CD4 lymphocyte act	24.7	Coronery artery SMC rest	40.3
CD45RO CD4 lymphocyte act	38.4	Coronery artery SMC TNFalpha + IL-1beta	31.6
CD8 lymphocyte act	13.1	Astrocytes rest	8.8
Secondary CD8 lymphocyte rest	13.9	Astrocytes TNFalpha + IL-1beta	7.5
Secondary CD8 lymphocyte act	6.2	KU-812 (Basophil) rest	36.9
CD4 lymphocyte none	2.7	KU-812 (Basophil) PMA/ionomycin	63.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	6.0	CCD1106 (Keratinocytes) none	48.0
LAK cells rest	18.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	14.9
LAK cells IL-2	18.7	Liver cirrhosis	7.4
LAK cells IL-2+IL-12	1.4	NCI-H292 none	45.4
LAK cells IL-2+IFN gamma	7.4	NCI-H292 IL-4	49.3
LAK cells IL-2+ IL-18	6.0	NCI-H292 IL-9	**100.0**
LAK cells PMA/ionomycin	20.4	NCI-H292 IL-13	43.5
NK Cells IL-2 rest	47.6	NCI-H292 IFN gamma	19.8
Two Way MLR 3 day	22.2	HPAEC none	9.2
Two Way MLR 5 day	6.8	HPAEC TNF alpha + IL-1 beta	43.8
Two Way MLR 7 day	7.5	Lung fibroblast none	47.6
PBMC rest	5.0	Lung fibroblast TNF alpha + IL-1 beta	53.6

PBMC PWM	7.7	Lung fibroblast IL-4	26.4
PBMC PHA-L	10.2	Lung fibroblast IL-9	32.8
Ramos (B cell) none	9.3	Lung fibroblast IL-13	15.4
Ramos (B cell) ionomycin	40.9	Lung fibroblast IFN gamma	9.5
B lymphocytes PWM	11.3	Dermal fibroblast CCD1070 rest	42.0
B lymphocytes CD40L and IL-4	27.7	Dermal fibroblast CCD1070 TNF alpha	70.7
EOL-1 dbcAMP	32.3	Dermal fibroblast CCD1070 IL-1 beta	24.7
EOL-1 dbcAMP PMA/ionomycin	3.7	Dermal fibroblast IFN gamma	24.8
Dendritic cells none	32.3	Dermal fibroblast IL-4	37.9
Dendritic cells LPS	9.8	Dermal Fibroblasts rest	35.4
Dendritic cells anti-CD40	10.7	Neutrophils TNFa+LPS	2.8
Monocytes rest	13.7	Neutrophils rest	27.2
Monocytes LPS	36.3	Colon	6.4
Macrophages rest	14.2	Lung	4.7
Macrophages LPS	17.4	Thymus	4.9
HUVEC none	26.4	Kidney	36.9
HUVEC starved	26.4		

CNS_neurodegeneration_v1.0 Summary: Ag7568 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene appears to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

Panel 4.1D Summary: Ag7568 Highest expression of this gene is seen in IL-9 treated NCI-H292 cells (CT=31.2). In addition, this gene is also expressed at moderate to low levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies,

inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

Y. CG149680-01 and CG149680-02: PROSTATE CANCER OVEREXPRESSED GENE 1

Expression of gene CG149680-02 and variant CG149680-01 was assessed using the primer-probe sets Ag4870 and Ag5280, described in Tables YA and YB. Results of the RTQ-PCR runs are shown in Tables YC, YD, YE and YF. Please note that Ag5280 is specific to CG149680-02.

Table YA. Probe Name Ag4870

Primers		Length	Start Position	SEQ ID No
Forward	5'-gcctgccttatctttctgaact-3'	22	707	314
Probe	TET-5'-ctttcctgcccctgaggaagtcaatt-3'-TAMRA	26	754	315
Reverse	5'-cactcagcttgatcttcttcgt-3'	22	782	316

Table YB. Probe Name Ag5280

Primers		Length	Start Position	SEQ ID No
Forward	5'-gctccctgttgatcattctg-3'	20	147	317
Probe	TET-5'-aacgagggcttctattccagcacgt-3'-TAMRA	25	170	318
Reverse	5'-cagcacatgacaccaggaa-3'	19	204	319

Table YC. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag5280, Run 234222214	Rel. Exp.(%) Ag5280, Run 237378555	issue Name	Rel. Exp.(%) Ag5280, Run 234222214	Rel. Exp.(%) Ag5280, Run 237378555
110967 COPD-F	17.2	17.6	112427 Match Control Psoriasis-F	100.0	100.0
110980 COPD-F	14.8	24.1	112418 Psoriasis-M	14.5	18.2
110968 COPD-M	10.5	25.3	112723 Match Control Psoriasis-M	5.8	5.6
110977 COPD-M	39.0	65.5	112419 Psoriasis-M	19.5	20.6
110989 Emphysema-F	43.2	33.0	112424 Match Control Psoriasis-M	6.8	3.6

110992 Emphysema-F	10.5	20.0	112420 Psoriasis-M	56.3	40.9
110993 Emphysema-F	14.5	27.2	112425 Match Control Psoriasis-M	62.4	47.3
110994 Emphysema-F	12.2	20.0	104689 (MF) OA Bone-Backus	11.2	13.7
110995 Emphysema-F	20.7	29.5	104690 (MF) Adj "Normal" Bone-Backus	11.5	8.8
110996 Emphysema-F	5.6	11.3	104691 (MF) OA Synovium-Backus	5.9	14.0
110997 Asthma-M	0.0	0.0	104692 (BA) OA Cartilage-Backus	0.0	0.0
111001 Asthma-F	13.8	22.5	104694 (BA) OA Bone-Backus	14.7	14.3
111002 Asthma-F	23.0	22.4	104695 (BA) Adj "Normal" Bone-Backus	14.8	15.9
111003 Atopic Asthma-F	19.6	20.4	104696 (BA) OA Synovium-Backus	6.3	13.2
111004 Atopic Asthma-F	30.4	69.7	104700 (SS) OA Bone-Backus	28.5	20.4
111005 Atopic Asthma-F	20.6	23.5	104701 (SS) Adj "Normal" Bone-Backus	23.7	13.0
111006 Atopic Asthma-F	2.6	7.9	104702 (SS) OA Synovium-Backus	16.6	20.9
111417 Allergy-M	22.8	26.8	117093 OA Cartilage Rep7	22.1	16.3
112347 Allergy-M	0.0	0.0	112672 OA Bone5	31.0	39.8
112349 Normal Lung-F	0.0	0.0	112673 OA Synovium5	16.5	9.3
112357 Normal Lung-F	59.5	79.0	112674 OA Synovial Fluid cells5	11.8	11.3
112354 Normal Lung-M	17.6	13.6	117100 OA Cartilage Rep14	2.6	5.7
112374 Crohns-F	23.5	2.9	112756 OA Bone9	95.3	81.8
112389 Match Control Crohns-F	10.3	2.8	112757 OA Synovium9	11.9	11.6
112375 Crohns-F	19.6	12.8	112758 OA Synovial Fluid Cells9	13.8	5.3
112732 Match Control Crohns-F	18.3	26.6	117125 RA Cartilage Rep2	11.7	25.0
112725 Crohns-M	2.5	3.0	113492 Bone2 RA	6.1	8.4
112387 Match Control Crohns-M	9.6	13.5	113493 Synovium2 RA	6.7	5.2
112378 Crohns-M	0.0	0.0	113494 Syn Fluid Cells RA	0.0	1.7

112390 Match Control Crohns-M	35.1	46.3	113499 Cartilage4 RA	1.6	10.6
112726 Crohns-M	36.9	26.2	113500 Bone4 RA	11.9	11.7
112731 Match Control Crohns-M	17.6	11.1	113501 Synovium4 RA	3.0	4.0
112380 Ulcer Col-F	19.9	24.0	113502 Syn Fluid Cells4 RA	0.0	0.0
112734 Match Control Ulcer Col-F	14.8	20.9	113495 Cartilage3 RA	7.8	2.2
112384 Ulcer Col-F	31.0	24.3	113496 Bone3 RA	20.6	9.0
112737 Match Control Ulcer Col-F	6.4	5.4	113497 Synovium3 RA	2.6	0.0
112386 Ulcer Col-F	2.7	6.1	113498 Syn Fluid Cells3 RA	5.4	10.5
112738 Match Control Ulcer Col-F	3.8	5.7	117106 Normal Cartilage Rep20	5.9	15.4
112381 Ulcer Col-M	0.0	0.0	113663 Bone3 Normal	0.0	0.0
112735 Match Control Ulcer Col-M	0.0	1.6	113664 Synovium3 Normal	0.0	0.0
112382 Ulcer Col-M	12.9	8.4	113665 Syn Fluid Cells3 Normal	0.0	0.0
112394 Match Control Ulcer Col-M	9.5	5.1	117107 Normal Cartilage Rep22	3.4	0.0
112383 Ulcer Col-M	9.4	17.2	113667 Bone4 Normal	13.4	9.5
112736 Match Control Ulcer Col-M	6.8	14.4	113668 Synovium4 Normal	11.2	4.1
112423 Psoriasis-F	5.2	0.0	113669 Syn Fluid Cells4 Normal	9.0	22.8

Table YD. General_screening_panel_v1.5

Tissue Name	**Rel. Exp.(%) Ag4870, Run 228903631**	**issue Name**	**Rel. Exp.(%) Ag4870, Run 228903631**
Adipose	2.0	Renal ca. TK-10	31.4
Melanoma* Hs688(A).T	6.0	Bladder	26.6
Melanoma* Hs688(B).T	4.2	Gastric ca. (liver met.) NCI-N87	7.4
Melanoma* M14	20.4	Gastric ca. KATO III	66.0

Melanoma* LOXIMVI	0.9	Colon ca. SW-948	10.2
Melanoma* SK-MEL-5	7.8	Colon ca. SW480	16.8
Squamous cell carcinoma SCC-4	0.5	Colon ca.* (SW480 met) SW620	63.7
Testis Pool	1.3	Colon ca. HT29	17.1
Prostate ca.* (bone met) PC-3	24.7	Colon ca. HCT-116	4.4
Prostate Pool	4.0	Colon ca. CaCo-2	36.6
Placenta	3.5	Colon cancer tissue	4.0
Uterus Pool	5.0	Colon ca. SW1116	5.7
Ovarian ca. OVCAR-3	0.8	Colon ca. Colo-205	6.7
Ovarian ca. SK-OV-3	1.2	Colon ca. SW-48	17.1
Ovarian ca. OVCAR-4	0.5	Colon Pool	6.3
Ovarian ca. OVCAR-5	17.1	Small Intestine Pool	5.0
Ovarian ca. IGROV-1	2.8	Stomach Pool	5.4
Ovarian ca. OVCAR-8	5.8	Bone Marrow Pool	2.3
Ovary	7.5	Fetal Heart	1.0
Breast ca. MCF-7	1.9	Heart Pool	3.3
Breast ca. MDA-MB-231	8.5	Lymph Node Pool	6.1
Breast ca. BT 549	3.2	Fetal Skeletal Muscle	2.5
Breast ca. T47D	0.4	Skeletal Muscle Pool	9.0
Breast ca. MDA-N	15.6	Spleen Pool	1.8
Breast Pool	8.0	Thymus Pool	4.2
Trachea	7.3	CNS cancer (glio/astro) U87-MG	1.6
Lung	1.5	CNS cancer (glio/astro) U-118-MG	0.9
Fetal Lung	11.6	CNS cancer (neuro;met) SK-N-AS	7.0
Lung ca. NCI-N417	1.6	CNS cancer (astro) SF-539	3.1
Lung ca. LX-1	49.3	CNS cancer (astro) SNB-75	14.9
Lung ca. NCI-H146	2.4	CNS cancer (glio) SNB-19	4.2
Lung ca. SHP-77	3.5	CNS cancer (glio) SF-295	10.0
Lung ca. A549	1.6	Brain (Amygdala) Pool	0.5
Lung ca. NCI-H526	0.2	Brain (cerebellum)	1.1
Lung ca. NCI-H23	0.8	Brain (fetal)	1.0
Lung ca. NCI-H460	1.2	Brain (Hippocampus) Pool	0.4
Lung ca. HOP-62	5.3	Cerebral Cortex Pool	0.4
Lung ca. NCI-H522	6.5	Brain (Substantia nigra) Pool	0.3
Liver	21.6	Brain (Thalamus) Pool	0.6
Fetal Liver	**100.0**	Brain (whole)	3.0
Liver ca. HepG2	68.3	Spinal Cord Pool	0.4
Kidney Pool	9.3	Adrenal Gland	8.1
Fetal Kidney	1.8	Pituitary gland Pool	0.3
Renal ca. 786-0	0.5	Salivary Gland	9.3
Renal ca. A498	0.1	Thyroid (female)	1.7
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	1.4
Renal ca. UO-31	2.3	Pancreas Pool	16.6

Table YE. Oncology cell line screening panel v3.1

Tissue Name	Rel. Exp.(%) Ag4870, Run 225053014	Tissu Name	Rel. Exp.(%) Ag4870, Run 225053014
Daoy Medulloblastoma/Cerebellum	1.8	Ca Ski_Cervical epidermoid carcinoma (metastasis)	2.1
TE671 Medulloblastom/Cerebellum	2.6	ES-2_Ovarian clear cell carcinoma	13.9
D283 Med Medulloblastoma/Cerebellum	11.2	Ramos/6h stim_ Stimulated with PMA/ionomycin 6h	94.6
PFSK-1 Primitive Neuroectodermal/Cerebellum	9.8	Ramos/14h stim_ Stimulated with PMA/ionomycin 14h	27.7
XF-498_CNS	3.4	MEG-01_Chronic myelogenous leukemia (megokaryoblast)	14.5
SNB-78_CNS/glioma	2.7	Raji_Burkitt's lymphoma	11.7
SF-268_CNS/glioblastoma	0.9	Daudi_Burkitt's lymphoma	27.5
T98G_Glioblastoma	5.7	U266_B-cell plasmacytoma/myeloma	0.9
SK-N-SH_Neuroblastoma (metastasis)	3.5	CA46_Burkitt's lymphoma	24.3
SF-295_CNS/glioblastoma	1.1	RL_non-Hodgkin's B-cell lymphoma	32.8
Cerebellum	0.3	JM1_pre-B-cell lymphoma/leukemia	3.5
Cerebellum	0.5	Jurkat_T cell leukemia	14.5
NCI-H292_Mucoepidermoid lung ca.	0.3	TF-1_Erythroleukemia	37.6
DMS-114_Small cell lung cancer	0.6	HUT 78_T-cell lymphoma	3.8
DMS-79_Small cell lung cancer/neuroendocrine	1.4	U937_Histiocytic lymphoma	48.6
NCI-H146_Small cell lung cancer/neuroendocrine	8.4	KU-812_Myelogenous leukemia	**100.0**
NCI-H526_Small cell lung cancer/neuroendocrine	1.7	769-P_Clear cell renal ca.	0.4
NCI-N417_Small cell lung cancer/neuroendocrine	5.1	Caki-2_Clear cell renal ca.	0.0
NCI-H82_Small cell lung cancer/neuroendocrine	10.1	SW 839_Clear cell renal ca.	1.1
NCI-H157_Squamous cell lung cancer (metastasis)	2.0	G401_Wilms' tumor	5.4
NCI-H1155_Large cell lung cancer/neuroendocrine	13.8	Hs766T_Pancreatic ca. (LN metastasis)	10.5
NCI-H1299_Large cell lung cancer/neuroendocrine	1.2	CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)	4.3
NCI-H727_Lung carcinoid	23.7	SU86.86_Pancreatic carcinoma (liver metastasis)	6.5
NCI-UMC-11_Lung carcinoid	52.1	BxPC-3_Pancreatic adenocarcinoma	0.7
LX-1_Small cell lung cancer	13.1	HPAC_Pancreatic adenocarcinoma	1.5

Colo-205_Colon cancer	13.0	MIA PaCa-2_Pancreatic ca.	0.0
KM12_Colon cancer	2.5	CFPAC-1_Pancreatic ductal adenocarcinoma	2.2
KM20L2_Colon cancer	10.8	PANC-1_Pancreatic epithelioid ductal ca.	0.3
NCI-H716_Colon cancer	21.0	T24_Bladder ca. (transitional cell)	6.5
SW-48_Colon adenocarcinoma	44.1	5637_Bladder ca.	1.7
SW1116_Colon adenocarcinoma	8.7	HT-1197_Bladder ca.	0.2
LS 174T_Colon adenocarcinoma	10.1	UM-UC-3_Bladder ca. (transitional cell)	0.4
SW-948_Colon adenocarcinoma	21.3	A204_Rhabdomyosarcoma	0.3
SW-480_Colon adenocarcinoma	12.9	HT-1080_Fibrosarcoma	9.3
NCI-SNU-5_Gastric ca.	12.3	MG-63_Osteosarcoma (bone)	1.7
KATO III_Stomach	43.2	SK-LMS-1_Leiomyosarcoma (vulva)	6.6
NCI-SNU-16_Gastric ca.	4.1	SJRH30_Rhabdomyosarcoma (met to bone marrow)	4.6
NCI-SNU-1_Gastric ca.	49.0	A431_Epidermoid ca.	0.4
RF-1_Gastric adenocarcinoma	6.6	WM266-4_Melanoma	4.1
RF-48_Gastric adenocarcinoma	6.0	DU 145_Prostate	2.0
MKN-45_Gastric ca.	21.5	MDA-MB-468_Breast adenocarcinoma	0.8
NCI-N87_Gastric ca.	2.7	SSC-4_Tongue	0.3
OVCAR-5_Ovarian ca.	0.7	SSC-9_Tongue	0.2
RL95-2_Uterine carcinoma	0.3	SSC-15_Tongue	1.5
HelaS3_Cervical adenocarcinoma	0.1	CAL 27_Squamous cell ca. of tongue	0.4

Table YF. Panel 4.1D

Tissue Name	Rel. Ep.(%) Ag5280, Run 230500483	Tissue Name	Rel. Exp.(%) Ag5280, Run 230500483
Secondary Th1 act	28.5	HUVEC IL-1beta	12.6
Secondary Th2 act	26.1	HUVEC IFN gamma	2.2
Secondary Tr1 act	11.0	HUVEC TNF alpha + IFN gamma	6.1
Secondary Th1 rest	5.7	HUVEC TNF alpha + IL4	12.7
Secondary Th2 rest	2.7	HUVEC IL-11	6.0
Secondary Tr1 rest	4.0	Lung Microvascular EC none	14.9
Primary Th1 act	2.9	Lung Microvascular EC TNFalpha + IL-1beta	5.3
Primary Th2 act	25.2	Microvascular Dermal EC none	0.0
Primary Tr1 act	8.2	Microsvasular Dermal EC TNFalpha + IL-1beta	5.1

Primary Th1 rest	1.8	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	5.4	Coronery artery SMC rest	1.5
CD45RO CD4 lymphocyte act	2.4	Coronery artery SMC TNFalpha + IL-1beta	3.6
CD8 lymphocyte act	4.2	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	8.3	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	3.0	KU-812 (Basophil) rest	**100.0**
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	90.8
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	7.6	Liver cirrhosis	9.8
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	3.8	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	3.3	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	8.4	NCI-H292 IL-13	2.9
NK Cells IL-2 rest	9.9	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.9	HPAEC none	4.8
Two Way MLR 5 day	3.8	HPAEC TNF alpha + IL-1 beta	4.5
Two Way MLR 7 day	2.1	Lung fibroblast none	27.4
PBMC rest	2.6	Lung fibroblast TNF alpha + IL-1 beta	7.6
PBMC PWM	0.0	Lung fibroblast IL-4	1.0
PBMC PHA-L	7.1	Lung fibroblast IL-9	7.1
Ramos (B cell) none	21.8	Lung fibroblast IL-13	7.2
Ramos (B cell) ionomycin	85.9	Lung fibroblast IFN gamma	14.7
B lymphocytes PWM	8.3	Dermal fibroblast CCD1070 rest	3.1
B lymphocytes CD40L and IL-4	28.5	Dermal fibroblast CCD1070 TNF alpha	19.5
EOL-1 dbcAMP	20.2	Dermal fibroblast CCD1070 IL-1 beta	9.9
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	9.4
Dendritic cells none	3.5	Dermal fibroblast IL-4	13.8
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	15.3
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	1.5
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	2.6
Macrophages LPS	0.0	Thymus	0.0

HUVEC none	9.9	Kidney	2.7
HUVEC starved	7.0		

AI_comprehensive panel_v1.0 Summary: Ag5280 Two experiments with the same probe and primer produce results that are in excellent agreement. Highest expression is in a sample derived from normal tissue adjacent to psoriasis (CTs=33). Low levels of expression are also seen in an osteoarthritic bone sample.

CNS_neurodegeneration_v1.0 Summary: Ag5280 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

General_screening_panel_v1.5 Summary: Ag4870 Highest expression of this gene, a PB39 homolog, is seen in the fetal liver (CT=25.6). Significant levels of expression are also seen in cell lines derived from lung, gastric, colon, renal, liver, ovarian, breast, prostate, melanoma and brain cancers. This expression in proliferetive samples suggests a role for this gene in cell proliferation and growth. This is consistent with data that shows to be upregulated in prostate cancer and tissues undergoing growth and differentiation. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of these cancers.

References:

Cole KA, Chuaqui RF, Katz K, Pack S, Zhuang Z, Cole CE, Lyne JC, Linehan WM, Liotta LA, Emmert-Buck MR. cDNA sequencing and analysis of POV1 (PB39): a novel gene up-regulated in prostate cancer. Genomics 1998 Jul 15;51(2):282-7

Stuart RO, Pavlova A, Beier D, Li Z, Krijanovski Y, Nigam SK. EEG1, a putative transporter expressed during epithelial organogenesis: comparison with embryonic transporter expression during nephrogenesis. Am J Physiol Renal Physiol 2001 Dec;281(6):F1148-56

Oncology_cell_line_screening_panel_v3.1 Summary: Ag4870 Highest expression of this gene is seen in a myelogenous leukemia cell line (CT=27.2). Moderate levels of expression are seen in other cell line samples on this panel, including samples from colon, gastric, and lung cancers, leukemias, and lymphomas. Please see Panel 1.5 for discussion of utility of this gene in cancer.

Panel 4.1D Summary: Ag5280 Prominent expression is seen in two samples derived from the basophil cell line KU-812 (CTs=32.5). Basophils release histamines and other biological modifiers in reponse to allergens and play an important role in the pathology of asthma and hypersensitivity reactions. Therefore, therapeutics designed against the putative protein encoded by this gene may reduce or inhibit inflammation by blocking basophil function in these diseases. In addition, these cells are a reasonable model for the inflammatory cells that take part in various inflammatory lung and bowel diseases, such as asthma, Crohn's disease, and ulcerative colitis. Therefore, expression of this gene could be used to differentiate between these samples and other samples on this panel adn as a marker of these cells. Furthermore, therapeutics that modulate the function of this gene product may reduce or eliminate the symptoms of patients suffering from asthma, Crohn's disease, and ulcerative colitis.

Z. CG149777-02: CYSTATIN D PRECURSOR

Expression of full-length physical clone CG149777-02 was assessed using the primer-probe set Ag6903, described in Table ZA.

Table ZA. Probe Name Ag6903

Primers		Length	Start Position	SEQ ID No
Forward	5'-ccacagacctcaatgacaagag-3'	22	110	320
Probe	TET-5'-cctggactttgccttcaatgaccag-3'-TAMRA	25	144	321
Reverse	5'-gaactcttcctctttcagttttgg-3'	24	169	322

General_screening_panel_v1.6 Summary: Ag6903 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

AA. CG150005-01: Glutamate binding protien

Expression of gene CG150005-01 was assessed using the primer-probe set Ag5633, described in Table AAA.

Table AAA. Probe Name Ag5633

Primers		Length	Start Position	SEQ ID No
Forward	5'-ccacctcctgtctactcattgt-3'	22	1341	323

Probe	TET-5'-catgagccctgtctgccagcttc-3'-TAMRA	23	1365	324
Reverse	5'-gctcaatccttggacctgtt-3'	20	1412	325

AI_comprehensive panel_v1.0 Summary: Ag5633 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

CNS_neurodegeneration_v1.0 Summary: Ag5633 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

General_screening_panel_v1.5 Summary: Ag5633 The amp plot indicates that there were experimental difficulties with this run; therefore, no conclusions can be drawn from this data. (Data not shown)

Panel 4.1D Summary: Ag5633 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

Panel 5D Summary: Ag5633 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

Panel CNS_1.1 Summary: Ag5633 Expression of this gene is low/undetectable in all samples on this panel (CTs>35).

AB. CG150189-01: Acetyl LDL Receptor

Expression of gene CG150189-01 was assessed using the primer-probe sets Ag3183 and Ag372, described in Tables ABA and ABB. Results of the RTQ-PCR runs are shown in Tables ABC, ABD, ABE, ABF, ABG and ABH.

Table ABA. Probe Name Ag3183

Primers		Length	Start Position	SEQ ID No
Forward	5'-aaggggacgagtgtgggatt-3'	20	212	326
Probe	TET-5'-tggcaccgaagtagccgtggcg-3'-TAMRA	22	301	327
Reverse	5'-gcgggcacttggtgtcgca-3'	19	325	328

Table ABB. Probe Name Ag372

Primers		Length	Start Position	SEQ ID No
Forward	5'-tgtaaccatgtcaccggcaa-3'	20	574	329

Probe	TET-5'-cgatccagcccgcgttgca-3'-TAMRA	19	604	330
Reverse	5'-ctcgccgtaagtgccattg-3'	19	648	331

Table ABC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3183, Run 216861424	issue Name	Rel. Exp.(%) Ag3183, Run 216861424
Adipose	5.2	Renal ca. TK-10	2.2
Melanoma* Hs688(A).T	82.9	Bladder	10.5
Melanoma* Hs688(B).T	**100.0**	Gastric ca. (liver met.) NCI-N87	2.0
Melanoma* M14	67.8	Gastric ca. KATO III	1.7
Melanoma* LOXIMVI	4.1	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	9.8	Colon ca. SW480	0.8
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.5
Testis Pool	2.7	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	3.7	Colon ca. HCT-116	0.5
Prostate Pool	3.0	Colon ca. CaCo-2	0.0
Placenta	6.4	Colon cancer tissue	19.8
Uterus Pool	5.0	Colon ca. SW1116	0.8
Ovarian ca. OVCAR-3	3.6	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	5.3	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	6.7	Colon Pool	9.7
Ovarian ca. OVCAR-5	4.9	Small Intestine Pool	3.7
Ovarian ca. IGROV-1	37.1	Stomach Pool	8.5
Ovarian ca. OVCAR-8	84.1	Bone Marrow Pool	1.3
Ovary	18.3	Fetal Heart	2.4
Breast ca. MCF-7	0.0	Heart Pool	5.1
Breast ca. MDA-MB-231	8.2	Lymph Node Pool	12.1
Breast ca. BT 549	14.5	Fetal Skeletal Muscle	2.4
Breast ca. T47D	15.5	Skeletal Muscle Pool	2.1
Breast ca. MDA-N	15.6	Spleen Pool	5.4
Breast Pool	10.7	Thymus Pool	12.2
Trachea	8.2	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.8	CNS cancer (glio/astro) U-118-MG	21.8
Fetal Lung	18.2	CNS cancer (neuro;met) SK-N-AS	34.2
Lung ca. NCI-N417	1.0	CNS cancer (astro) SF-539	90.8
Lung ca. LX-1	0.9	CNS cancer (astro) SNB-75	71.7
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	31.2
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	82.9
Lung ca. A549	1.4	Brain (Amygdala) Pool	3.5
Lung ca. NCI-H526	6.3	Brain (cerebellum)	0.0

Lung ca. NCI-H23	0.0	Brain (fetal)	2.0
Lung ca. NCI-H460	0.5	Brain (Hippocampus) Pool	1.4
Lung ca. HOP-62	7.2	Cerebral Cortex Pool	1.0
Lung ca. NCI-H522	7.3	Brain (Substantia nigra) Pool	6.2
Liver	2.1	Brain (Thalamus) Pool	1.1
Fetal Liver	2.1	Brain (whole)	1.7
Liver ca. HepG2	9.9	Spinal Cord Pool	6.8
Kidney Pool	24.5	Adrenal Gland	1.3
Fetal Kidney	18.8	Pituitary gland Pool	1.1
Renal ca. 786-0	1.8	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	4.2
Renal ca. ACHN	7.3	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	7.3	Pancreas Pool	15.6

Table ABD. Panel 1

Tissue Name	Rel. Ex.(%) Ag372, Run 98747566	Tissue Name	Rel. Exp.(%) Ag372, Run 98747566
Endothelial cells	13.1	Renal ca. 786-0	14.3
Endothelial cells (treated)	6.1	Renal ca. A498	4.4
Pancreas	9.7	Renal ca. RXF 393	14.3
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	15.7
Adrenal gland	27.0	Renal ca. UO-31	8.5
Thyroid	31.0	Renal ca. TK-10	0.4
Salivary gland	9.7	Liver	26.4
Pituitary gland	41.2	Liver (fetal)	11.8
Brain (fetal)	13.0	Liver ca. (hepatoblast) HepG2	25.3
Brain (whole)	11.0	Lung	41.8
Brain (amygdala)	15.1	Lung (fetal)	32.5
Brain (cerebellum)	9.7	Lung ca. (small cell) LX-1	2.6
Brain (hippocampus)	12.9	Lung ca. (small cell) NCI-H69	8.2
Brain (substantia nigra)	10.2	Lung ca. (s.cell var.) SHP-77	2.9
Brain (thalamus)	11.8	Lung ca. (large cell)NCI-H460	22.5
Brain (hypothalamus)	42.9	Lung ca. (non-sm. cell) A549	7.4
Spinal cord	15.7	Lung ca. (non-s.cell) NCI-H23	4.3
glio/astro U87-MG	1.0	Lung ca. (non-s.cell) HOP-62	29.5
glio/astro U-118-MG	12.2	Lung ca. (non-s.cl) NCI-H522	44.8
astrocytoma SW1783	25.9	Lung ca. (squam.) SW 900	10.3
neuro*; met SK-N-AS	66.4	Lung ca. (squam.) NCI-H596	4.7
astrocytoma SF-539	56.3	Mammary gland	41.5
astrocytoma SNB-75	23.3	Breast ca.* (pl.ef) MCF-7	3.5

glioma SNB-19	23.5	Breast ca.* (pl.ef) MDA-MB-231	6.1
glioma U251	18.2	Breast ca.* (pl. ef) T47D	14.6
glioma SF-295	48.3	Breast ca. BT-549	4.0
Heart	32.1	Breast ca. MDA-N	32.8
Skeletal muscle	16.0	Ovary	67.8
Bone marrow	7.0	Ovarian ca. OVCAR-3	13.3
Thymus	11.3	Ovarian ca. OVCAR-4	9.4
Spleen	25.2	Ovarian ca. OVCAR-5	4.0
Lymph node	15.4	Ovarian ca. OVCAR-8	**100.0**
Colon (ascending)	5.9	Ovarian ca. IGROV-1	21.9
Stomach	10.2	Ovarian ca. (ascites) SK-OV-3	6.3
Small intestine	62.9	Uterus	35.8
Colon ca. SW480	0.0	Placenta	33.9
Colon ca.* SW620 (SW480 met)	2.1	Prostate	31.4
Colon ca. HT29	0.2	Prostate ca.* (bone met) PC-3	23.7
Colon ca. HCT-116	2.4	Testis	16.5
Colon ca. CaCo-2	0.3	Melanoma Hs688(A).T	44.4
Colon ca. HCT-15	5.9	Melanoma* (met) Hs688(B).T	55.9
Colon ca. HCC-2998	12.1	Melanoma UACC-62	70.2
Gastric ca. * (liver met) NCI-N87	6.4	Melanoma M14	45.1
Bladder	51.4	Melanoma LOX IMVI	6.5
Trachea	21.0	Melanoma* (met) SK-MEL-5	27.2
Kidney	32.8	Melanoma SK-MEL-28	0.0
Kidney (fetal)	67.8		

Table ABE. Panel 1.3D

Tissue Name	Rel. Exp.(% Ag3183, Run 167927219	Tissue Name	Rel. Exp.(%) Ag3183, Run 167927219
Liver adenocarcinoma	5.5	Kidney (fetal)	**100.0**
Pancreas	0.0	Renal ca. 786-0	5.1
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	16.6
Adrenal gland	4.7	Renal ca. RXF 393	9.9
Thyroid	2.8	Renal ca. ACHN	17.8
Salivary gland	6.1	Renal ca. UO-31	1.8
Pituitary gland	1.6	Renal ca. TK-10	0.0
Brain (fetal)	0.9	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	9.9
Brain (amygdala)	1.7	Liver ca. (hepatoblast) HepG2	7.3
Brain (cerebellum)	0.0	Lung	7.4
Brain (hippocampus)	0.0	Lung (fetal)	11.2

Brain (substantia nigra)	1.1	Lung ca. (small cell) LX-1	1.7
Brain (thalamus)	2.5	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	5.1
Spinal cord	3.5	Lung ca. (large cell)NCI-H460	2.9
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	7.7	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	14.4	Lung ca. (non-s.cell) HOP-62	6.0
neuro*; met SK-N-AS	7.5	Lung ca. (non-s.cl) NCI-H522	5.2
astrocytoma SF-539	52.1	Lung ca. (squam.) SW 900	5.1
astrocytoma SNB-75	49.7	Lung ca. (squam.) NCI-H596	1.0
glioma SNB-19	3.1	Mammary gland	10.2
glioma U251	13.9	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	25.9	Breast ca.* (pl.ef) MDA-MB-231	3.7
Heart (fetal)	14.4	Breast ca.* (pl.ef) T47D	11.8
Heart	7.0	Breast ca. BT-549	1.5
Skeletal muscle (fetal)	10.7	Breast ca. MDA-N	5.9
Skeletal muscle	0.0	Ovary	13.0
Bone marrow	1.7	Ovarian ca. OVCAR-3	1.5
Thymus	3.3	Ovarian ca. OVCAR-4	1.8
Spleen	6.9	Ovarian ca. OVCAR-5	0.6
Lymph node	9.2	Ovarian ca. OVCAR-8	5.3
Colorectal	1.6	Ovarian ca. IGROV-1	0.0
Stomach	4.7	Ovarian ca.* (ascites) SK-OV-3	2.2
Small intestine	11.7	Uterus	27.9
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.8
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	4.6
Colon ca. HCT-116	0.0	Testis	1.7
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	50.3
Colon ca. tissue(ODO3866)	15.2	Melanoma* (met) Hs688(B).T	84.7
Colon ca. HCC-2998	0.0	Melanoma UACC-62	42.3
Gastric ca.* (liver met) NCI-N87	0.9	Melanoma M14	13.9
Bladder	2.0	Melanoma LOX IMVI	10.3
Trachea	5.8	Melanoma* (met) SK-MEL-5	0.0
Kidney	22.8	Adipose	9.0

Table ABF. Panel 4D

Tissue Name	Rel. Ep.(%) Ag3183, Run 164317572	Tissue Name	Rel. Exp.(%) Ag3183, Run 164317572
Secondary Th1 act	0.0	HUVEC IL-1beta	5.7

Secondary Th2 act	0.0	HUVEC IFN gamma	18.4
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	9.6
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	11.2
Secondary Th2 rest	0.0	HUVEC IL-11	19.3
Secondary Tr1 rest	2.5	Lung Microvascular EC none	21.6
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	32.1
Primary Th2 act	0.0	Microvascular Dermal EC none	10.7
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	5.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	11.7
Primary Th2 rest	2.3	Small airway epithelium none	5.6
Primary Tr1 rest	0.7	Small airway epithelium TNFalpha + IL-1beta	4.9
CD45RA CD4 lymphocyte act	29.9	Coronery artery SMC rest	47.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	45.4
CD8 lymphocyte act	0.0	Astrocytes rest	82.4
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	75.3
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.4	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.1	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	5.6
LAK cells IL-2+IL-12	0.0	Lupus kidney	9.3
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	2.5
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	3.8
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	32.1
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	22.4
PBMC rest	0.0	Lung fibroblast none	44.4
PBMC PWM	1.8	Lung fibroblast TNF alpha + IL-1 beta	28.3
PBMC PHA-L	0.0	Lung fibroblast IL-4	62.9
Ramos (B cell) none	0.0	Lung fibroblast IL-9	**100.0**
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	73.7
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	57.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	91.4
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	31.2

EOL-1 dbcAMP PMA/ionomycin	1.2	Dermal fibroblast CCD1070 IL-1 beta	69.7
Dendritic cells none	1.8	Dermal fibroblast IFN gamma	18.3
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	22.4
Dendritic cells anti-CD40	0.4	IBD Colitis 2	0.0
Monocytes rest	2.8	IBD Crohn's	0.8
Monocytes LPS	0.0	Colon	8.9
Macrophages rest	0.0	Lung	18.6
Macrophages LPS	0.0	Thymus	9.5
HUVEC none	8.4	Kidney	3.0
HUVEC starved	29.9		

Table ABG. Panel 5D

Tissue Name	**Rel. Exp.(%) Ag313, Run 172171149**	**Tissue Name**	**Rel. Exp.(%) Ag3183, Run 172171149**
97457_Patient-02go_adipose	11.3	94709_Donor 2 AM - A_adipose	31.6
97476_Patient-07sk_skeletal muscle	21.3	94710_Donor 2 AM - B_adipose	47.0
97477_Patient-07ut_uterus	3.3	94711_Donor 2 AM - C_adipose	36.3
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	33.0
97481_Patient-08sk_skeletal muscle	9.0	94713_Donor 2 AD - B_adipose	27.0
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	13.5
97483_Patient-08pl_placenta	5.4	94742_Donor 3 U - A_Mesenchymal Stem Cells	27.9
97486_Patient-09sk_skeletal muscle	2.4	94743_Donor 3 U - B_Mesenchymal Stem Cells	**100.0**
97487_Patient-09ut_uterus	11.3	94730_Donor 3 AM - A_adipose	69.3
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	58.6
97492_Patient-10ut_uterus	3.4	94732_Donor 3 AM - C_adipose	11.2
97493_Patient-10pl_placenta	8.4	94733_Donor 3 AD - A_adipose	47.0
97495_Patient-11go_adipose	8.1	94734_Donor 3 AD - B_adipose	31.6
97496_Patient-11sk_skeletal muscle	0.0	94735_Donor 3 AD - C_adipose	44.8
97497_Patient-11ut_uterus	17.4	77138_Liver_HepG2untreated	9.1
97498_Patient-11pl_placenta	0.0	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	15.7	81735_Small Intestine	6.3
97501_Patient-12sk_skeletal muscle	0.0	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	13.8	82685_Small intestine_Duodenum	0.0

97503_Patient-12pl_placenta	4.8	90650_Adrenal_Adrenocortical adenoma	15.4
94721_Donor 2 U - A_Mesenchymal Stem Cells	31.2	72410_Kidney_HRCE	14.2
94722_Donor 2 U - B_Mesenchymal Stem Cells	33.7	72411_Kidney_HRE	7.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	20.7	73139_Uterus_Uterine smooth muscle cells	11.3

Table ABH. general oncology screening panel_v_2.4

Tissue Name	Rel. Exp.(%) Ag3183, Run 259733268	Tissue Nme	Rel. Exp.(%) Ag3183, Run 259733268
Colon cancer 1	24.7	Bladder NAT 2	0.0
Colon NAT 1	19.2	Bladder NAT 3	0.0
Colon cancer 2	6.6	Bladder NAT 4	15.9
Colon NAT 2	0.0	Prostate adenocarcinoma 1	3.2
Colon cancer 3	23.8	Prostate adenocarcinoma 2	0.0
Colon NAT 3	10.4	Prostate adenocarcinoma 3	4.5
Colon malignant cancer 4	3.1	Prostate adenocarcinoma 4	16.7
Colon NAT 4	0.0	Prostate NAT 5	14.7
Lung cancer 1	21.5	Prostate adenocarcinoma 6	0.0
Lung NAT 1	0.0	Prostate adenocarcinoma 7	7.6
Lung cancer 2	**100.0**	Prostate adenocarcinoma 8	0.0
Lung NAT 2	0.0	Prostate adenocarcinoma 9	5.3
Squamous cell carcinoma 3	39.2	Prostate NAT 10	0.0
Lung NAT 3	6.3	Kidney cancer 1	9.8
Metastatic melanoma 1	11.1	Kidney NAT 1	6.1
Melanoma 2	5.8	Kidney cancer 2	15.0
Melanoma 3	0.0	Kidney NAT 2	18.9
Metastatic melanoma 4	61.1	Kidney cancer 3	11.1
Metastatic melanoma 5	49.7	Kidney NAT 3	7.5
Bladder cancer 1	0.0	Kidney cancer 4	6.9
Bladder NAT 1	0.0	Kidney NAT 4	8.6
Bladder cancer 2	4.3		

General_screening_panel_v1.4 Summary: Ag3183 Highest expression of this gene is seen in a melanoma cell line (CT=31.5). Prominent expression is seen in a cluster of cell lines derived ovarian, melanoma, and brain cells. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a

marker of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of ovarian, melanoma and brain cancers.

Panel 1 Summary: Ag4337 Highest expression of this gene is detected in a ovarian cancer OVCAR-8 cell line (CT=26.5). High to Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from liver, gastric, colon, lung, renal, breast, ovarian, melanoma and brain cancers. Therefore, therapeutic modulation of this gene may be useful in the treatment of these cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, and adult and fetal liver. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 1.3D Summary: Ag3183 Highest expression of this gene is seen in fetal kidney (CT=32.2). In addition, prominent expression is seen in clusters of cell lines derived from melanoma, and brain cancer cell lines. Please see Panel 1 for discussion of utility of this gene in cancer. In another experiment (run 167966980) the amp plot indicates that there were experimental difficulties with this run; therefore, no conclusions can be drawn from this data (Data not shown).

Panel 4D Summary: Ag3183 Highest expression of this gene is detected in activated lung fibroblast (CT=31.9). This gene is also expressed in resting and treated fibroblasts, endothelium, and epithelium and activated naive T cells (CD4+ CD45RA cells). Interestingly, this gene is up-regulated activated in naive T cells (CD4+ CD45RA cells; CT=33.6) as compared to resting CD4 cells (CT=40). Furthermore, in activated memory T cells (CD45RO CD4 lymphocyte) or CD4 Th1 or Th2 cells (CTs>37), the expression of this gene is strongly down regulated suggesting a role for this putative protein in differentiation or activation of naive T cells. Activated T cells then initiate the inflammatory process by secreting cytokines and chemokines, activating B cells and

inducing B cell antibody production, and inducing the extravasation of leukocytes including other T cells into inflammatory sites. Therefore, therapeutics that inhibit the action of this gene product may block T cell activation in response to tissue transplant and reduce or block rejection. These therapeutic drugs may also reduce or prevent inflammation in asthma/allergy, psoriasis, arthritis and diabetes in which activated T cells play a pivotal role. Expression of this gene may also serve as a diagnostic or experimental tools to identify naive activated T cells and discriminate them from more differentiated activated T cells (memory T cells).

References:

Study of LDL and acetylated LDL endocytosis by mononuclear cells in HIV infection. Juompan L, Puel J, Fournie GJ, Benoist H Biochim Biophys Acta 1995 Aug 15;1272(1):21-8.

Panel 5D Summary: Ag3182 Highest expression of this gene is seen in a sample of mesenchymal stem cells (CT=34.2). Low but significant levels of expression are also seen in adipose tissue, in agreement with expression in Panel 1. Please see Panel 1 for discussion of this gene in metabolic disease.

general oncology screening panel_v_2.4 Summary: Ag3183 Expression is seen in a lung cancer sample (CT=34.9). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of lung cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of lung cancer.

AC. CG150267-01: Type Ia membrane protein

Expression of gene CG150267-01 was assessed using the primer-probe set Ag7560, described in Table ACA. Results of the RTQ-PCR runs are shown in Tables ACB and ACC.

Table ACA. Probe Name Ag7560

Primers		Length	Start Position	SEQ ID No
Forward	5'-gcacctgcttcggatatttt-3'	20	560	332
Probe	TET-5'-tttccctctgtacttatgccgccagt-3'-TAMRA	26	586	333
Reverse	5'-ggagccggttcaaatcatac-3'	20	617	334

Table ACB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7560, Run 308750602	issue Name	Rel. Exp.(%) Ag7560, Run 308750602
AD 1 Hippo	8.8	Control (Path) 3 Temporal Ctx	6.7
AD 2 Hippo	12.3	Control (Path) 4 Temporal Ctx	26.4
AD 3 Hippo	10.7	AD 1 Occipital Ctx	22.5
AD 4 Hippo	8.2	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	51.4	AD 3 Occipital Ctx	8.1
AD 6 Hippo	24.7	AD 4 Occipital Ctx	26.4
Control 2 Hippo	25.0	AD 5 Occipital Ctx	29.5
Control 4 Hippo	9.8	AD 6 Occipital Ctx	13.4
Control (Path) 3 Hippo	1.1	Control 1 Occipital Ctx	5.3
AD 1 Temporal Ctx	21.2	Control 2 Occipital Ctx	51.4
AD 2 Temporal Ctx	23.2	Control 3 Occipital Ctx	4.5
AD 3 Temporal Ctx	6.1	Control 4 Occipital Ctx	7.9
AD 4 Temporal Ctx	19.2	Control (Path) 1 Occipital Ctx	62.4
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	9.7
AD 5 Sup Temporal Ctx	34.2	Control (Path) 3 Occipital Ctx	1.1
AD 6 Inf Temporal Ctx	18.6	Control (Path) 4 Occipital Ctx	11.1
AD 6 Sup Temporal Ctx	19.2	Control 1 Parietal Ctx	5.9
Control 1 Temporal Ctx	3.2	Control 2 Parietal Ctx	19.5
Control 2 Temporal Ctx	19.1	Control 3 Parietal Ctx	17.6
Control 3 Temporal Ctx	9.5	Control (Path) 1 Parietal Ctx	71.2
Control 3 Temporal Ctx	5.5	Control (Path) 2 Parietal Ctx	17.0
Control (Path) 1 Temporal Ctx	48.3	Control (Path) 3 Parietal Ctx	5.0
Control (Path) 2 Temporal Ctx	17.6	Control (Path) 4 Parietal Ctx	22.2

Table ACC. Panel 4.1D

Tissue Name	Rel. Exp (%) Ag7560, Run 308748085	Tissue Name	Rel. Exp.(%) Ag7560, Run 308748085
Secondary Th1 act	0.0	HUVEC IL-1beta	3.2
Secondary Th2 act	0.0	HUVEC IFN gamma	5.2
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	6.5
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	3.8
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0

Primary Th2 act	0.0	Microvascular Dermal EC none	20.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	3.5
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	4.6
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	2.9
CD8 lymphocyte act	0.0	Astrocytes rest	75.3
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	23.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	5.6
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	6.1
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	1.8
LAK cells IL-2+IL-12	0.0	NCI-H292 none	10.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	3.8
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	6.1
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	4.2
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	4.9	Lung fibroblast none	3.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	10.3
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	16.3
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	3.5
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	4.2
EOL-1 dbcAMP	11.2	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	6.2	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	3.8	Dermal fibroblast IL-4	4.2
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0

Monocytes LPS	0.0	Colon	4.0
Macrophages rest	6.5	Lung	3.8
Macrophages LPS	0.0	Thymus	10.5
HUVEC none	0.0	Kidney	**100.0**
HUVEC starved	7.5		

CNS_neurodegeneration_v1.0 Summary: Ag7560 No differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. However, this panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag7560 Highest expression of this gene is detected in kidney (CT=33.8). Therefore, expression of this gene could be used to differentiate the kidney-derived sample from other samples on this panel and as a marker of kidney tissue. In addition, therapeutic targeting of the expression or function of this gene may modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

Low but significant levels of expression of this gene is also seen in resting astrocytes. Therefore, therapeutic modulation of this gene or the encoded protein could be important in the treatment of multiple sclerosis or other inflammatory diseases of the CNS.

AD. CG150362-01: OTOFERLIN

Expression of gene CG150362-01 was assessed using the primer-probe set Ag5684, described in Table ADA. Results of the RTQ-PCR runs are shown in Table ADB.

Table ADA. Probe Name Ag5684

Primers		Length	Start Position	SEQ ID No
Forward	5'-cctggtatttgagcagttgatc-3'	22	3187	335
Probe	TET-5'-atcactaatggaggctcctcctgcag-3'-TAMRA	26	3230	336
Reverse	5'-gccaaacttattgtggtcaaat-3'	22	3265	337

Table ADB. Panel 4.1D

Tissue Name	Rel. Exp (%) Ag5684, Run 246498693	Tissue Name	Rel. Exp.(%) Ag5684, Run 246498693
Secondary Th1 act	0.0	HUVEC IL-1beta	1.9
Secondary Th2 act	0.0	HUVEC IFN gamma	15.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	2.6
Secondary Tr1 rest	0.0	Lung Microvascular EC none	17.3
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.1
Primary Th2 act	1.9	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	6.4
Primary Th2 rest	1.7	Small airway epithelium none	1.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	5.3
CD45RA CD4 lymphocyte act	2.0	Coronery artery SMC rest	13.7
CD45RO CD4 lymphocyte act	1.6	Coronery artery SMC TNFalpha + IL-1beta	25.3
CD8 lymphocyte act	2.3	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.6	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	2.5	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.4	CCD1106 (Keratinocytes) none	7.2
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	5.7
LAK cells IL-2	0.0	Liver cirrhosis	6.4
LAK cells IL-2+IL-12	0.0	NCI-H292 none	66.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	57.8
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	92.7
LAK cells PMA/ionomycin	1.7	NCI-H292 IL-13	100.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	51.8
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	3.0
Two Way MLR 7 day	0.5	Lung fibroblast none	6.3
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	3.8

PBMC PWM	0.0	Lung fibroblast IL-4	1.5
PBMC PHA-L	2.3	Lung fibroblast IL-9	3.5
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	16.0	Lung fibroblast IFN gamma	2.1
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.3
B lymphocytes CD40L and IL-4	2.3	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	1.5
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	1.9
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	4.5
HUVEC starved	2.9		

CNS_neurodegeneration_v1.0 Summary: Ag5684 Expression of this gene is low/undetectable (CTs > 34.8) across all of the samples on this panel (data not shown).

General_screening_panel_v1.5 Summary: Ag5684 The amp plot indicates that there were experimental difficulties with this run; therefore, no conclusions can be drawn from this data. (Data not shown).

Panel 4.1D Summary: Ag5684 Highest expression of this gene is detected in IL-13 treated NCI-H292 cell line (CT=30.4). This gene is also expressed in a cluster of treated and untreated NCI-H292 cell line, a human airway epithelial cell line that produces mucins. Mucus overproduction is an important feature of bronchial asthma and chronic obstructive pulmonary disease samples. This gene is also expressed at lower but still significant levels in ionomycin treated Ramos B cells, activated HUVEC cells, activated bronchial epithelium and small airway epithelium, resting lung fibroblasts, coronery artery SMC and keratinocytes. Therefore, therapeutics designed with the protein encoded by this gene may reduce or eliminate symptoms caused by inflammation in lung epithelia in chronic obstructive pulmonary disease, asthma, allergy, and emphysema.

AE. CG150637-02: T-CELL SURFACE GLYCOPROTEIN CD1B PRECURSOR

Expression of full-length physical clone CG150637-02 was assessed using the primer-probe set Ag7126, described in Table AEA. Results of the RTQ-PCR runs are shown in Table AEB.

Table AEA. Probe Name Ag7126

Primers		Length	Start Position	SEQ ID No
Forward	5'-ggatgcggggaaacc-3'	15	718	338
Probe	TET-5'-acctccattggctcaattgttttggc -3'-TAMRA	26	735	339
Reverse	5'-ccataatgcaaggcatagca-3'	20	787	340

Table AEB. Panel 4.1D

Tissue Name	Rel. Exp.(% Ag7126, Run 306518354	Tissue Name	Rel. Exp.(%) Ag7126, Run 306518354
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0

LAK cells rest	12.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	11.2	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	1.1	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	**100.0**	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	64.6	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	86.5	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	19.5
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag7126 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag7126 Highest expression of this gene is detected in dendritic cells (CT=32). Moderate to low levels of expression of this gene is restricted to resting and activated dendritic cells, and thymus. Dendritic cells are powerful antigen-presenting cells (APC), whose function is pivotal in the initiation and maintenance of normal immune responses. Autoimmunity and inflammation may be reduced by suppression of this function. Therefore, therapeutic modulation of the protein encoded by

this gene may be important in the treatment of autoimmune and inflammatory diseases such as Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

AF. CG150694-01: Microfibril-associated glycoprotein 2 precursor

Expression of full-length physical clone CG150694-01 was assessed using the primer-probe set Ag7144, described in Table AFA. Results of the RTQ-PCR runs are shown in Tables AFB and AFC.

Table AFA. Probe Name Ag7144

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-gatgaaacagagtgctgggat-3'	21	166	341
Probe	TET-5'-atttacctgcacaaggctctactctgtgc-3'-TAMRA	29	192	342
Reverse	5'-actgatgaatgcattgtttaacc-3'	23	228	343

Table AFB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7144, Run 306518753	issue Name	Rel. Exp.(%) Ag7144, Run 306518753
AD 1 Hippo	19.1	Control (Path) 3 Temporal Ctx	13.1
AD 2 Hippo	28.1	Control (Path) 4 Temporal Ctx	2.6
AD 3 Hippo	3.1	AD 1 Occipital Ctx	4.0
AD 4 Hippo	6.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	14.1	AD 3 Occipital Ctx	5.3
AD 6 Hippo	**100.0**	AD 4 Occipital Ctx	1.6
Control 2 Hippo	3.0	AD 5 Occipital Ctx	7.4
Control 4 Hippo	20.2	AD 6 Occipital Ctx	7.0
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	9.2
AD 1 Temporal Ctx	2.9	Control 2 Occipital Ctx	15.0
AD 2 Temporal Ctx	15.7	Control 3 Occipital Ctx	0.0
AD 3 Temporal Ctx	0.0	Control 4 Occipital Ctx	3.2
AD 4 Temporal Ctx	12.0	Control (Path) 1 Occipital Ctx	31.0
AD 5 Inf Temporal Ctx	51.1	Control (Path) 2 Occipital Ctx	5.3
AD 5 Sup Temporal Ctx	39.5	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	22.2	Control (Path) 4 Occipital Ctx	8.0
AD 6 Sup Temporal Ctx	23.8	Control 1 Parietal Ctx	0.0

Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	27.5
Control 2 Temporal Ctx	1.0	Control 3 Parietal Ctx	2.3
Control 3 Temporal Ctx	0.0	Control (Path) 1 Parietal Ctx	17.9
Control 3 Temporal Ctx	1.1	Control (Path) 2 Parietal Ctx	15.1
Control (Path) 1 Temporal Ctx	5.3	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	4.7	Control (Path) 4 Parietal Ctx	8.7

Table AFC. Panel 4.1D

Tissue Name	**Rel. Ex.(%) Ag7144, Run 306518356**	**Tissue Name**	**Rel. Exp.(%) Ag7144, Run 306518356**
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.1	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.3
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.3
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	2.3
Primary Th2 rest	0.0	Small airway epithelium none	0.7
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	2.5
CD45RA CD4 lymphocyte act	0.8	Coronery artery SMC rest	91.4
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	**100.0**
CD8 lymphocyte act	0.0	Astrocytes rest	24.7
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	55.9
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	3.2
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.4
LAK cells IL-2	0.0	Liver cirrhosis	0.4
LAK cells IL-2+IL-12	0.0	NCI-H292 none	13.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	25.2
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	39.2

LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	41.5
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	11.8
Two Way MLR 3 day	0.0	HPAEC none	0.1
Two Way MLR 5 day	0.1	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	5.7
B lymphocytes CD40L and IL-4	0.1	Dermal fibroblast CCD1070 TNF alpha	4.3
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	3.3
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	3.9
Dendritic cells none	0.0	Dermal fibroblast IL-4	5.6
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	10.3
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.7	Colon	1.3
Macrophages rest	0.0	Lung	0.2
Macrophages LPS	0.0	Thymus	1.4
HUVEC none	0.0	Kidney	0.2
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag7144 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

Panel 4.1D Summary: Ag7144 Highest expression of this gene is detected in resting and activated coronery artery SMC (CTs=28). Moderate levels of expression of this gene is also seen in astrocytes, keratinocytes, mucoepidermoid NCI-H292 cells, activated bronchial and small airway epithelius and dermal fibroblasts. In addition, low levels of expression of this gene are also seen in colon and thymus. Therefore, therapeutic modulation of this gene or its protein product through the use of antibody or small molecule drug may be useful in the treatment of autoimmune and inflammatory diseases

such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, osteoarthritis, multiple sclerosis and other inflammatory diseases of the CNS.

AG. CG151069-01: membrane protein AK027056.1

Expression of gene CG151069-01 was assessed using the primer-probe set Ag7562, described in Table AGA. Results of the RTQ-PCR runs are shown in Tables AGB and AGC.

Table AGA. Probe Name Ag7562

Primers		Length	Start Position	SEQ ID No
Forward	5'-aatctgtggctggggtcat-3'	19	861	344
Probe	TET-5'-cccctggacgtctccgtcacaat-3'-TAMRA	23	887	345
Reverse	5'-cactcattgtgaaaataggctgata-3'	25	923	346

Table AGB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7562, Run 308750605	issue Name	Rel. Exp.(%) Ag7562, Run 308750605
AD 1 Hippo	12.9	Control (Path) 3 Temporal Ctx	0.9
AD 2 Hippo	17.8	Control (Path) 4 Temporal Ctx	9.2
AD 3 Hippo	9.4	AD 1 Occipital Ctx	24.8
AD 4 Hippo	11.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	40.6	AD 3 Occipital Ctx	9.4
AD 6 Hippo	31.4	AD 4 Occipital Ctx	27.5
Control 2 Hippo	15.7	AD 5 Occipital Ctx	15.8
Control 4 Hippo	3.8	AD 6 Occipital Ctx	15.3
Control (Path) 3 Hippo	3.0	Control 1 Occipital Ctx	3.9
AD 1 Temporal Ctx	38.4	Control 2 Occipital Ctx	19.3
AD 2 Temporal Ctx	23.0	Control 3 Occipital Ctx	7.2
AD 3 Temporal Ctx	8.9	Control 4 Occipital Ctx	9.2
AD 4 Temporal Ctx	23.3	Control (Path) 1 Occipital Ctx	70.7
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	11.5
AD 5 Sup Temporal Ctx	35.6	Control (Path) 3 Occipital Ctx	1.4
AD 6 Inf Temporal Ctx	53.6	Control (Path) 4 Occipital Ctx	7.7
AD 6 Sup Temporal Ctx	37.9	Control 1 Parietal Ctx	9.1
Control 1 Temporal Ctx	1.2	Control 2 Parietal Ctx	52.5

Control 2 Temporal Ctx	14.3	Control 3 Parietal Ctx	9.0
Control 3 Temporal Ctx	7.9	Control (Path) 1 Parietal Ctx	31.9
Control 3 Temporal Ctx	6.1	Control (Path) 2 Parietal Ctx	17.2
Control (Path) 1 Temporal Ctx	30.8	Control (Path) 3 Parietal Ctx	1.3
Control (Path) 2 Temporal Ctx	10.2	Control (Path) 4 Parietal Ctx	17.4

Table AGC. Panel 4.1D

Tissue Name	Rel. Exp (%) Ag7562, Run 308748090	Tissue Name	Rel. Exp.(%) Ag7562, Run 308748090
Secondary Th1 act	0.0	HUVEC IL-1beta	4.2
Secondary Th2 act	0.0	HUVEC IFN gamma	17.2
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	1.2
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	1.2
Secondary Th2 rest	0.0	HUVEC IL-11	3.2
Secondary Tr1 rest	0.0	Lung Microvascular EC none	88.3
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	17.6
Primary Th2 act	0.0	Microvascular Dermal EC none	11.0
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	8.7
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.8
Primary Tr1 rest	0.7	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	1.8	Coronery artery SMC rest	1.6
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	1.6
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.7
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	57.8
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	92.7
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.5	Liver cirrhosis	3.7
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	0.0

NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	69.7
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	100.0
Two Way MLR 7 day	0.0	Lung fibroblast none	40.6
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	6.4
PBMC PWM	0.0	Lung fibroblast IL-4	17.1
PBMC PHA-L	0.0	Lung fibroblast IL-9	10.3
Ramos (B cell) none	0.0	Lung fibroblast IL-13	6.6
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	16.2
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.6	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	78.5	Dermal fibroblast CCD1070 IL-1 beta	1.7
EOL-1 dbcAMP PMA/ionomycin	40.1	Dermal fibroblast IFN gamma	1.4
Dendritic cells none	0.0	Dermal fibroblast IL-4	5.1
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	18.2
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	1.8
Macrophages rest	0.0	Lung	5.8
Macrophages LPS	0.0	Thymus	1.2
HUVEC none	9.7	Kidney	3.3
HUVEC starved	7.1		

CNS_neurodegeneration_v1.0 Summary: Ag7562 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

Panel 4.1D Summary: Ag7562 Highest expression of this gene is detected in alpha + IL-1 beta treated HPAEC (CT=32.2). Moderate to low levels of expression of this gene is also seen in eosinophils, lung microvascular endothelial cells, basophils, HPAEC, and activated lung fibroblasts. Therefore, therapeutic modulation of this gene or its protein product through the use of small molecule drug or antibodies may be useful in the treatment of autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

AH. CG151189-01: Type IIIb membrane protein

Expression of gene CG151189-01 was assessed using the primer-probe set Ag7561, described in Table AHA. Results of the RTQ-PCR runs are shown in Tables AHB and AHC.

Table AHA. Probe Name Ag7561

Primers		Length	Start Position	SEQ ID No
Forward	5'-ctggagggcctgtcaaa-3'	17	440	347
Probe	TET-5'-cctccgatggcgaaaccagcatt-3'-TAMRA	23	486	348
Reverse	5'-tcacagaatttagtaagcgttgg-3'	23	524	349

Table AHB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7561, Run 308750603	issue Name	Rel. Exp.(%) Ag7561, Run 308750603
AD 1 Hippo	14.7	Control (Path) 3 Temporal Ctx	4.5
AD 2 Hippo	32.5	Control (Path) 4 Temporal Ctx	36.3
AD 3 Hippo	9.7	AD 1 Occipital Ctx	17.3
AD 4 Hippo	7.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	84.1	AD 3 Occipital Ctx	7.8
AD 6 Hippo	68.8	AD 4 Occipital Ctx	21.6
Control 2 Hippo	26.8	AD 5 Occipital Ctx	19.5
Control 4 Hippo	12.9	AD 6 Occipital Ctx	37.1
Control (Path) 3 Hippo	7.9	Control 1 Occipital Ctx	5.1
AD 1 Temporal Ctx	20.2	Control 2 Occipital Ctx	51.1
AD 2 Temporal Ctx	35.6	Control 3 Occipital Ctx	18.9
AD 3 Temporal Ctx	6.4	Control 4 Occipital Ctx	10.0
AD 4 Temporal Ctx	27.5	Control (Path) 1 Occipital Ctx	97.3
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	15.3
AD 5 SupTemporal Ctx	54.0	Control (Path) 3 Occipital Ctx	3.6
AD 6 Inf Temporal Ctx	73.2	Control (Path) 4 Occipital Ctx	19.6
AD 6 Sup Temporal Ctx	74.2	Control 1 Parietal Ctx	9.9
Control 1 Temporal Ctx	6.9	Control 2 Parietal Ctx	55.9
Control 2 Temporal Ctx	33.4	Control 3 Parietal Ctx	21.5
Control 3 Temporal Ctx	14.8	Control (Path) 1 Parietal Ctx	63.7
Control 4 Temporal Ctx	13.9	Control (Path) 2 Parietal Ctx	29.7
Control (Path) 1 Temporal Ctx	57.8	Control (Path) 3 Parietal Ctx	5.0
Control (Path) 2 Temporal Ctx	44.8	Control (Path) 4 Parietal Ctx	34.6

Table AHC. Panel 4.1D

Tissue Name	Rel. Exp.(% Ag7561, Run 308748088	Tissue Name	Rel. Exp.(%) Ag7561, Run 308748088
Secondary Th1 act	28.3	HUVEC IL-1beta	19.2
Secondary Th2 act	100.0	HUVEC IFN gamma	15.3
Secondary Tr1 act	25.0	HUVEC TNF alpha + IFN gamma	5.6
Secondary Th1 rest	1.9	HUVEC TNF alpha + IL4	9.6
Secondary Th2 rest	4.2	HUVEC IL-11	6.1
Secondary Tr1 rest	3.8	Lung Microvascular EC none	40.1
Primary Th1 act	4.7	Lung Microvascular EC TNFalpha + IL-1beta	10.1
Primary Th2 act	26.1	Microvascular Dermal EC none	4.7
Primary Tr1 act	18.0	Microsvasular Dermal EC TNFalpha + IL-1beta	9.0
Primary Th1 rest	2.3	Bronchial epithelium TNFalpha + IL1beta	3.7
Primary Th2 rest	2.4	Small airway epithelium none	2.8
Primary Tr1 rest	0.8	Small airway epithelium TNFalpha + IL-1beta	14.0
CD45RA CD4 lymphocyte act	27.0	Coronery artery SMC rest	14.5
CD45RO CD4 lymphocyte act	28.1	Coronery artery SMC TNFalpha + IL-1beta	12.4
CD8 lymphocyte act	10.0	Astrocytes rest	14.8
Secondary CD8 lymphocyte rest	7.2	Astrocytes TNFalpha + IL-1beta	12.4
Secondary CD8 lymphocyte act	6.7	KU-812 (Basophil) rest	4.6
CD4 lymphocyte none	2.7	KU-812 (Basophil) PMA/ionomycin	7.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	6.0	CCD1106 (Keratinocytes) none	22.2
LAK cells rest	11.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.1
LAK cells IL-2	6.9	Liver cirrhosis	2.3
LAK cells IL-2+IL-12	0.7	NCI-H292 none	8.6
LAK cells IL-2+IFN gamma	5.4	NCI-H292 IL-4	12.5
LAK cells IL-2+ IL-18	4.9	NCI-H292 IL-9	15.7
LAK cells PMA/ionomycin	18.9	NCI-H292 IL-13	15.8
NK Cells IL-2 rest	26.2	NCI-H292 IFN gamma	7.0
Two Way MLR 3 day	9.0	HPAEC none	4.5
Two Way MLR 5 day	4.0	HPAEC TNF alpha + IL-1 beta	24.3
Two Way MLR 7 day	3.6	Lung fibroblast none	13.6

PBMC rest	1.4	Lung fibroblast TNF alpha + IL-1 beta	18.8
PBMC PWM	6.7	Lung fibroblast IL-4	13.9
PBMC PHA-L	8.0	Lung fibroblast IL-9	31.4
Ramos (B cell) none	6.3	Lung fibroblast IL-13	11.1
Ramos (B cell) ionomycin	14.9	Lung fibroblast IFN gamma	26.8
B lymphocytes PWM	6.3	Dermal fibroblast CCD1070 rest	27.2
B lymphocytes CD40L and IL-4	20.7	Dermal fibroblast CCD1070 TNF alpha	45.7
EOL-1 dbcAMP	14.4	Dermal fibroblast CCD1070 IL-1 beta	20.7
EOL-1 dbcAMP PMA/ionomycin	14.1	Dermal fibroblast IFN gamma	6.1
Dendritic cells none	13.4	Dermal fibroblast IL-4	16.4
Dendritic cells LPS	5.5	Dermal Fibroblasts rest	12.2
Dendritic cells anti-CD40	6.7	Neutrophils TNFa+LPS	3.5
Monocytes rest	3.3	Neutrophils rest	11.7
Monocytes LPS	22.4	Colon	2.0
Macrophages rest	8.3	Lung	1.4
Macrophages LPS	6.5	Thymus	4.6
HUVEC none	12.7	Kidney	10.7
HUVEC starved	17.9		

CNS_neurodegeneration_v1.0 Summary: Ag7561 No differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. However, this panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag7561 Highest expression of this gene is detected in activated secondary Th2 cells (CT=29.3). This gene is expressed at moderate to low levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions

associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

AI. CG151801-01: Occludin like membrane protein

Expression of gene CG151801-01 was assessed using the primer-probe set Ag7563, described in Table AIA. Results of the RTQ-PCR runs are shown in Table AIB.

Table AIA. Probe Name Ag7563

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-actttctcacataaagcaaagaattc-3'	26	1629	350
Probe	TET-5'-ccttgtacatcccaattcattacttta tca-3'-TAMRA	30	1662	351
Reverse	5'-gtggtttcaaataagcgttaagaat-3'	25	1694	352

Table AIB. Panel 4.1D

Tissue Name	Rel. Ex.(%) Ag7563, Run 308748092	Tissue Name	Rel. Exp.(%) Ag7563, Run 308748092
Secondary Th1 act	0.0	HUVEC IL-1beta	2.8
Secondary Th2 act	0.0	HUVEC IFN gamma	3.8
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	20.9
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.8
Primary Th2 act	0.0	Microvascular Dermal EC none	2.5
Primary Tr1 act	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	12.2
Primary Th2 rest	0.0	Small airway epithelium none	18.4
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	**100.0**
CD45RA CD4 lymphocyte act	0.0	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0

CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	41.5
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	33.4
LAK cells IL-2	0.0	Liver cirrhosis	4.3
LAK cells IL-2+IL-12	0.0	NCI-H292 none	31.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	57.4
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	45.7
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	59.9
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	21.6
Two Way MLR 3 day	2.7	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	25.9
Two Way MLR 7 day	0.0	Lung fibroblast none	2.8
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	3.7
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	3.5
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	7.1
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	2.4
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	29.3	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	6.1
Macrophages rest	0.0	Lung	2.3
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	1.9	Kidney	45.7
HUVEC starved	5.3		

CNS_neurodegeneration_v1.0 Summary: Ag7563 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag7563 Highest expression of this gene is seen in TNFalpha + IL-1beta treated small airway epithelium (CT=34). Therefore, expression of this gene may be used to distinguish activated small airway epithelium from other samples in this panel. In addition, low levels of expression of this gene are also seen in cytokine activated NCI-H292 cells, a human airway epithelial cell line that produces mucins. Therefore, modulation of the expression or activity of the protein encoded by this gene through the application of small molecule therapeutics or antibodies may be useful in the treatment of asthma, COPD, and emphysema.

AJ. CG165961-01 and CG165961-02: Secretory carrier-associated membrane protein 3

Expression of full-length physical clone CG165961-01 and variant CG165961-02 was assessed using the primer-probe set Ag7569, described in Table AJA. Results of the RTQ-PCR runs are shown in Tables AJB and AJC. Please note that CG165961-01 represents a full-length physical clone of the CG165961-02 gene, validating the prediction of the gene sequence.

Table AJA. Probe Name Ag7569

Primers		Length	Start Position	SEQ ID No
Forward	5'-ctggctcttctcctgaacttc-3'	21	594	353
Probe	TET-5'-ccagcttctgtgtggaaaccaacaat-3'-TAMRA	26	555	354
Reverse	5'-aggacccagaggatagaaagc-3'	21	520	355

Table AJB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7569, Run 308751132	issue Name	Rel. Exp.(%) Ag7569, Run 308751132
AD 1 Hippo	22.8	Control (Path) 3 Temporal Ctx	7.2
AD 2 Hippo	41.5	Control (Path) 4 Temporal Ctx	34.2
AD 3 Hippo	11.3	AD 1 Occipital Ctx	21.8
AD 4 Hippo	9.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	88.3	AD 3 Occipital Ctx	10.8
AD 6 Hippo	68.8	AD 4 Occipital Ctx	25.7
Control 2 Hippo	35.4	AD 5 Occipital Ctx	27.9

Control 4 Hippo	19.8	AD 6 Occipital Ctx	49.3
Control (Path) 3 Hippo	8.1	Control 1 Occipital Ctx	6.9
AD 1 Temporal Ctx	21.5	Control 2 Occipital Ctx	86.5
AD 2 Temporal Ctx	37.6	Control 3 Occipital Ctx	18.7
AD 3 Temporal Ctx	8.4	Control 4 Occipital Ctx	9.9
AD 4 Temporal Ctx	21.0	Control (Path) 1 Occipital Ctx	88.3
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	12.8
AD 5 SupTemporal Ctx	54.0	Control (Path) 3 Occipital Ctx	7.0
AD 6 Inf Temporal Ctx	62.4	Control (Path) 4 Occipital Ctx	16.6
AD 6 Sup Temporal Ctx	57.0	Control 1 Parietal Ctx	9.6
Control 1 Temporal Ctx	6.8	Control 2 Parietal Ctx	44.1
Control 2 Temporal Ctx	50.7	Control 3 Parietal Ctx	25.2
Control 3 Temporal Ctx	18.3	Control (Path) 1 Parietal Ctx	71.2
Control 4 Temporal Ctx	12.1	Control (Path) 2 Parietal Ctx	25.0
Control (Path) 1 Temporal Ctx	56.6	Control (Path) 3 Parietal Ctx	7.2
Control (Path) 2 Temporal Ctx	34.4	Control (Path) 4 Parietal Ctx	44.4

Table AJC. Panel 4.1D

Tissue Name	Rel. Ex.(%) Ag7569, Run 308748454	Tissue Name	Rel. Exp.(%) Ag7569, Run 308748454
Secondary Th1 act	80.1	HUVEC IL-1beta	74.2
Secondary Th2 act	72.7	HUVEC IFN gamma	42.9
Secondary Tr1 act	26.2	HUVEC TNF alpha + IFN gamma	24.5
Secondary Th1 rest	3.7	HUVEC TNF alpha + IL4	25.9
Secondary Th2 rest	2.9	HUVEC IL-11	21.5
Secondary Tr1 rest	4.5	Lung Microvascular EC none	72.7
Primary Th1 act	12.6	Lung Microvascular EC TNFalpha + IL-1beta	41.5
Primary Th2 act	59.5	Microvascular Dermal EC none	11.3
Primary Tr1 act	76.3	Microsvasular Dermal EC TNFalpha + IL-1beta	26.8
Primary Th1 rest	5.3	Bronchial epithelium TNFalpha + IL1beta	22.8
Primary Th2 rest	7.0	Small airway epithelium none	16.6
Primary Tr1 rest	2.5	Small airway epithelium TNFalpha + IL-1beta	33.0
CD45RA CD4 lymphocyte act	55.9	Coronery artery SMC rest	32.1
CD45RO CD4 lymphocyte act	72.7	Coronery artery SMC TNFalpha + IL-1beta	39.0
CD8 lymphocyte act	31.2	Astrocytes rest	15.7
Secondary CD8 lymphocyte rest	34.9	Astrocytes TNFalpha + IL-1beta	16.0

Secondary CD8 lymphocyte act	15.1	KU-812 (Basophil) rest	31.0
CD4 lymphocyte none	4.5	KU-812 (Basophil) PMA/ionomycin	16.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	8.5	CCD1106 (Keratinocytes) none	44.4
LAK cells rest	14.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	15.7
LAK cells IL-2	16.4	Liver cirrhosis	6.5
LAK cells IL-2+IL-12	1.5	NCI-H292 none	34.4
LAK cells IL-2+IFN gamma	8.8	NCI-H292 IL-4	50.3
LAK cells IL-2+ IL-18	8.1	NCI-H292 IL-9	57.8
LAK cells PMA/ionomycin	33.0	NCI-H292 IL-13	44.1
NK Cells IL-2 rest	48.3	NCI-H292 IFN gamma	20.4
Two Way MLR 3 day	24.1	HPAEC none	9.7
Two Way MLR 5 day	11.8	HPAEC TNF alpha + IL-1 beta	68.8
Two Way MLR 7 day	12.5	Lung fibroblast none	48.6
PBMC rest	1.9	Lung fibroblast TNF alpha + IL-1 beta	63.3
PBMC PWM	25.7	Lung fibroblast IL-4	34.6
PBMC PHA-L	17.0	Lung fibroblast IL-9	59.5
Ramos (B cell) none	31.6	Lung fibroblast IL-13	23.3
Ramos (B cell) ionomycin	80.7	Lung fibroblast IFN gamma	99.3
B lymphocytes PWM	15.8	Dermal fibroblast CCD1070 rest	83.5
B lymphocytes CD40L and IL-4	44.8	Dermal fibroblast CCD1070 TNF alpha	**100.0**
EOL-1 dbcAMP	21.0	Dermal fibroblast CCD1070 IL-1 beta	47.6
EOL-1 dbcAMP PMA/ionomycin	3.3	Dermal fibroblast IFN gamma	32.8
Dendritic cells none	14.2	Dermal fibroblast IL-4	33.0
Dendritic cells LPS	13.8	Dermal Fibroblasts rest	47.3
Dendritic cells anti-CD40	8.0	Neutrophils TNFa+LPS	0.5
Monocytes rest	4.5	Neutrophils rest	1.4
Monocytes LPS	37.1	Colon	3.4
Macrophages rest	25.0	Lung	6.7
Macrophages LPS	21.8	Thymus	3.5
HUVEC none	43.5	Kidney	30.6
HUVEC starved	50.0		

CNS_neurodegeneration_v1.0 Summary: Ag7569 No differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. However, this panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central

nervous system disorders such as Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag7569 Highest expression of this gene is detected in TNF alpha treated dermal fibroblast (CT=29.9). This gene is expressed at moderate to low levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

AK. CG51595-03 and CG51595-06 and CG51595-07: Thrombospondin related protein

Expression of gene CG51595-06 and variants CG51595-03 and CG51595-07 was assessed using the primer-probe sets Ag815 and Ag127, described in Tables AKA and AKB. Results of the RTQ-PCR runs are shown in Tables AKC, AKD, AKE, AKF, AKG, AKH, AKI and AKJ. Please note that Ag127 is specific to CG51595-06 and CG51595-07 only.

Table AKA. Probe Name Ag815

Primers		Length	Start Position	SEQ ID No
Forward	5'-tgtgctcagcacatggtcta-3'	20	1716	356
Probe	TET-5'-acacctgctcagggaaaacgacagaa-3'-TAMRA	26	1754	357
Reverse	5'-tcgtgctcgtatctgtttcc-3'	20	1781	358

Table AKB. Probe Name Ag127

Primers	Sequence	Length	Start Position	SEQ ID No

Forward	5'-cctgccaggatgactgtcaatt-3'	22	2516	359
Probe	TET-5'-ccagctggtccaagttttcttcatgcaa-3'-TAMRA	28	2540	360
Reverse	5'-tggtcctaactgcaccacagtct-3'	23	2571	361

Table AKC. AI comprehensive panel v1.0

Tissue Name	Rel. Exp.(%) Ag815, Run 257809397	issue Name	Rel. Exp.(%) Ag815, Run 257809397
110967 COPD-F	14.0	112427 Match Control Psoriasis-F	33.2
110980 COPD-F	9.0	112418 Psoriasis-M	10.6
110968 COPD-M	7.5	112723 Match Control Psoriasis-M	1.1
110977 COPD-M	12.1	112419 Psoriasis-M	10.9
110989 Emphysema-F	18.7	112424 Match Control Psoriasis-M	12.4
110992 Emphysema-F	10.3	112420 Psoriasis-M	36.6
110993 Emphysema-F	10.2	112425 Match Control Psoriasis-M	27.7
110994 Emphysema-F	3.8	104689 (MF) OA Bone-Backus	18.2
110995 Emphysema-F	20.4	104690 (MF) Adj "Normal" Bone-Backus	9.3
110996 Emphysema-F	3.7	104691 (MF) OA Synovium-Backus	7.6
110997 Asthma-M	1.6	104692 (BA) OA Cartilage-Backus	4.5
111001 Asthma-F	9.9	104694 (BA) OA Bone-Backus	10.3
111002 Asthma-F	14.2	104695 (BA) Adj "Normal" Bone-Backus	12.8
111003 Atopic Asthma-F	31.4	104696 (BA) OA Synovium-Backus	8.8
111004 Atopic Asthma-F	1.3	104700 (SS) OA Bone-Backus	8.9
111005 Atopic Asthma-F	10.1	104701 (SS) Adj "Normal" Bone-Backus	9.0
111006 Atopic Asthma-F	1.2	104702 (SS) OA Synovium-Backus	17.3
111417 Allergy-M	7.6	117093 OA Cartilage Rep7	25.2
112347 Allergy-M	3.1	112672 OA Bone5	29.3
112349 Normal Lung-F	2.1	112673 OA Synovium5	11.8
112357 Normal Lung-F	6.9	112674 OA Synovial Fluid cells5	10.4
112354 Normal Lung-M	6.5	117100 OA Cartilage Rep14	2.0
112374 Crohns-F	8.0	112756 OA Bone9	6.7
112389 Match Control Crohns-F	7.0	112757 OA Synovium9	1.6
112375 Crohns-F	7.4	112758 OA Synovial Fluid Cells9	12.8
112732 Match Control Crohns-F	0.1	117125 RA Cartilage Rep2	14.6
112725 Crohns-M	18.3	113492 Bone2 RA	8.7
112387 Match Control Crohns-M	4.6	113493 Synovium2 RA	2.1
112378 Crohns-M	2.1	113494 Syn Fluid Cells RA	4.4

112390 Match Control Crohns-M	22.7	113499 Cartilage4 RA	6.6
112726 Crohns-M	31.2	113500 Bone4 RA	7.9
112731 Match Control Crohns-M	18.7	113501 Synovium4 RA	4.8
112380 Ulcer Col-F	14.9	113502 Syn Fluid Cells4 RA	3.9
112734 Match Control Ulcer Col-F	2.1	113495 Cartilage3 RA	4.8
112384 Ulcer Col-F	51.4	113496 Bone3 RA	9.0
112737 Match Control Ulcer Col-F	13.3	113497 Synovium3 RA	2.5
112386 Ulcer Col-F	3.6	113498 Syn Fluid Cells3 RA	9.3
112738 Match Control Ulcer Col-F	0.9	117106 Normal Cartilage Rep20	5.7
112381 Ulcer Col-M	0.8	113663 Bone3 Normal	8.8
112735 Match Control Ulcer Col-M	**100.0**	113664 Synovium3 Normal	0.9
112382 Ulcer Col-M	12.2	113665 Syn Fluid Cells3 Normal	4.0
112394 Match Control Ulcer Col-M	3.7	117107 Normal Cartilage Rep22	13.8
112383 Ulcer Col-M	36.3	113667 Bone4 Normal	7.2
112736 Match Control Ulcer Col-M	1.1	113668 Synovium4 Normal	15.6
112423 Psoriasis-F	12.9	113669 Syn Fluid Cells4 Normal	17.7

<u>Table AKD. Panel 1</u>

Tissue Name	**Rel. Ex.(%) Ag127, Run 87588501**	**Tissue Name**	**Rel. Exp.(%) Ag127, Run 87588501**
Endothelial cells	9.2	Renal ca. 786-0	0.0
Endothelial cells (treated)	2.2	Renal ca. A498	0.1
Pancreas	1.4	Renal ca. RXF 393	0.1
Pancreatic ca. CAPAN 2	0.4	Renal ca. ACHN	0.1
Adrenal gland	4.9	Renal ca. UO-31	0.6
Thyroid	4.8	Renal ca. TK-10	0.2
Salivary gland	0.7	Liver	2.6
Pituitary gland	4.2	Liver (fetal)	1.1
Brain (fetal)	7.1	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	33.9	Lung	4.7
Brain (amygdala)	6.0	Lung (fetal)	3.2
Brain (cerebellum)	47.6	Lung ca. (small cell) LX-1	0.0
Brain (hippocampus)	15.5	Lung ca. (small cell) NCI-H69	7.9
Brain (substantia nigra)	5.8	Lung ca. (s.cell var.) SHP-77	0.0

Brain (thalamus)	7.8	Lung ca. (large cell)NCI-H460	0.0
Brain (hypothalamus)	2.9	Lung ca. (non-sm. cell) A549	15.4
Spinal cord	6.6	Lung ca. (non-s.cell) NCI-H23	12.2
glio/astro U87-MG	1.2	Lung ca. (non-s.cell) HOP-62	1.9
glio/astro U-118-MG	1.3	Lung ca. (non-s.cl) NCI-H522	0.1
astrocytoma SW1783	0.7	Lung ca. (squam.) SW 900	7.1
neuro*; met SK-N-AS	18.6	Lung ca. (squam.) NCI-H596	8.4
astrocytoma SF-539	0.0	Mammary gland	12.0
astrocytoma SNB-75	0.4	Breast ca.* (pl.ef) MCF-7	0.0
glioma SNB-19	0.7	Breast ca.* (pl.ef) MDA-MB-231	0.0
glioma U251	3.7	Breast ca.* (pl. ef) T47D	0.9
glioma SF-295	0.3	Breast ca. BT-549	0.0
Heart	1.4	Breast ca. MDA-N	0.1
Skeletal muscle	0.1	Ovary	7.4
Bone marrow	0.2	Ovarian ca. OVCAR-3	1.0
Thymus	2.1	Ovarian ca. OVCAR-4	0.0
Spleen	2.1	Ovarian ca. OVCAR-5	6.0
Lymph node	1.1	Ovarian ca. OVCAR-8	2.9
Colon (ascending)	6.4	Ovarian ca. IGROV-1	4.9
Stomach	5.1	Ovarian ca. (ascites) SK-OV-3	0.3
Small intestine	1.6	Uterus	25.7
Colon ca. SW480	0.0	Placenta	**100.0**
Colon ca.* SW620 (SW480 met)	0.0	Prostate	4.0
Colon ca. HT29	0.0	Prostate ca.* (bone met) PC-3	0.0
Colon ca. HCT-116	0.0	Testis	33.7
Colon ca. CaCo-2	0.2	Melanoma Hs688(A).T	0.0
Colon ca. HCT-15	0.3	Melanoma* (met) Hs688(B).T	0.1
Colon ca. HCC-2998	1.5	Melanoma UACC-62	0.0
Gastric ca. * (liver met) NCI-N87	2.0	Melanoma M14	0.4
Bladder	11.3	Melanoma LOX IMVI	0.0
Trachea	2.4	Melanoma* (met) SK-MEL-5	15.9
Kidney	17.1	Melanoma SK-MEL-28	0.1
Kidney (fetal)	31.4		

Table AKE. Panel 1.2

Tissue Name	**Rel. Exp.(%) g815, Run 118424515**	**Rel. Exp.(%) Ag815, Run 122039235**	**Tissue Name**	**Rel. Exp.(%) Ag815, Run 118424515**	**Rel. Exp.(%) Ag815, Run 122039235**
Endothelial cells	94.6	17.9	Renal ca. 786-0	0.1	0.0
Heart (Fetal)	4.7	4.2	Renal ca. A498	0.1	0.0
Pancreas	5.4	0.4	Renal ca. RXF 393	0.0	0.0

Pancreatic ca. CAPAN 2	0.3	0.1	Renal ca. ACHN	0.1	0.0
Adrenal Gland	6.3	4.7	Renal ca. UO-31	0.4	0.3
Thyroid	9.0	1.0	Renal ca. TK-10	0.2	0.1
Salivary gland	2.1	1.4	Liver	3.6	1.7
Pituitary gland	20.3	4.6	Liver (fetal)	1.9	1.8
Brain (fetal)	31.4	5.9	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (whole)	19.6	16.8	Lung	3.1	2.9
Brain (amygdala)	4.5	5.9	Lung (fetal)	4.5	2.0
Brain (cerebellum)	6.8	10.3	Lung ca. (small cell) LX-1	0.0	0.0
Brain (hippocampus)	8.2	10.8	Lung ca. (small cell) NCI-H69	27.5	8.4
Brain (thalamus)	14.2	9.9	Lung ca. (s.cell var.) SHP-77	8.4	4.1
Cerebral Cortex	73.7	100.0	Lung ca. (large cell)NCI-H460	26.4	36.6
Spinal cord	7.5	5.7	Lung ca. (non-sm. cell) A549	68.3	36.9
glio/astro U87-MG	3.4	1.2	Lung ca. (non-s.cell) NCI-H23	21.3	28.1
glio/astro U-118-MG	2.0	0.7	Lung ca. (non-s.cell) HOP-62	7.4	2.3
astrocytoma SW1783	0.8	0.2	Lung ca. (non-s.cl) NCI-H522	0.4	0.0
neuro*; met SK-N-AS	88.9	15.2	Lung ca. (squam.) SW 900	8.2	9.4
astrocytoma SF-539	0.1	0.0	Lung ca. (squam.) NCI-H596	47.0	13.3
astrocytoma SNB-75	0.3	0.1	Mammary gland	4.3	2.2
glioma SNB-19	1.2	1.1	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma U251	12.8	7.9	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
glioma SF-295	0.8	0.2	Breast ca.* (pl. ef) T47D	1.1	0.8
Heart	8.0	5.8	Breast ca. BT-549	0.1	0.0
Skeletal Muscle	3.6	0.9	Breast ca. MDA-N	0.4	0.1
Bone marrow	0.4	0.2	Ovary	8.9	6.6
Thymus	0.4	0.4	Ovarian ca. OVCAR-3	3.5	0.8
Spleen	1.8	0.6	Ovarian ca. OVCAR-4	0.1	0.0
Lymph node	2.6	1.4	Ovarian ca. OVCAR-5	21.6	9.2

Colorectal Tissue	1.0	1.3	Ovarian ca. OVCAR-8	3.0	2.3
Stomach	2.8	3.2	Ovarian ca. IGROV-1	27.9	5.6
Small intestine	3.3	1.2	Ovarian ca. (ascites) SK-OV-3	1.8	1.1
Colon ca. SW480	0.0	0.0	Uterus	8.2	4.8
Colon ca.* SW620 (SW480 met)	0.0	0.0	Placenta	**100.0**	95.3
Colon ca. HT29	0.2	0.0	Prostate	3.5	3.0
Colon ca. HCT-116	0.0	0.0	Prostate ca.* (bone met) PC-3	0.2	0.1
Colon ca. CaCo-2	0.3	0.1	Testis	8.5	2.9
Colon ca. Tissue (ODO3866)	1.8	2.6	Melanoma Hs688(A).T	0.0	0.2
Colon ca. HCC-2998	6.2	1.8	Melanoma* (met) Hs688(B).T	0.1	0.1
Gastric ca.* (liver met) NCI-N87	3.5	2.0	Melanoma UACC-62	0.1	0.1
Bladder	21.6	12.5	Melanoma M14	0.3	0.1
Trachea	1.9	0.8	Melanoma LOX IMVI	0.0	0.0
Kidney	34.4	49.0	Melanoma* (met) SK-MEL-5	26.2	16.7
Kidney (fetal)	39.2	72.2			

<u>Table AKF. Panel 1.3D</u>

Tissue Name	**Rel. Exp.() Ag815, Run 152862062**	**Tissue Name**	**Rel. Exp.(%) Ag815, Run 152862062**
Liver adenocarcinoma	0.0	Kidney (fetal)	11.3
Pancreas	0.6	Renal ca. 786-0	0.1
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.7
Adrenal gland	1.4	Renal ca. RXF 393	0.1
Thyroid	0.9	Renal ca. ACHN	0.0
Salivary gland	0.2	Renal ca. UO-31	0.0
Pituitary gland	2.8	Renal ca. TK-10	0.2
Brain (fetal)	7.1	Liver	1.1
Brain (whole)	7.4	Liver (fetal)	0.5
Brain (amygdala)	4.3	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	3.4	Lung	2.0
Brain (hippocampus)	13.0	Lung (fetal)	5.0
Brain (substantia nigra)	1.1	Lung ca. (small cell) LX-1	5.8

Brain (thalamus)	8.0	Lung ca. (small cell) NCI-H69	23.8
Cerebral Cortex	100.0	Lung ca. (s.cell var.) SHP-77	9.0
Spinal cord	6.1	Lung ca. (large cell)NCI-H460	6.3
glio/astro U87-MG	0.9	Lung ca. (non-sm. cell) A549	13.4
glio/astro U-118-MG	3.8	Lung ca. (non-s.cell) NCI-H23	32.8
astrocytoma SW1783	0.4	Lung ca. (non-s.cell) HOP-62	2.1
neuro*; met SK-N-AS	57.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	3.3
astrocytoma SNB-75	3.3	Lung ca. (squam.) NCI-H596	12.1
glioma SNB-19	0.6	Mammary gland	0.6
glioma U251	8.5	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.8	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	4.1	Breast ca.* (pl.ef) T47D	0.5
Heart	1.3	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	28.7	Breast ca. MDA-N	0.0
Skeletal muscle	0.8	Ovary	13.8
Bone marrow	0.7	Ovarian ca. OVCAR-3	1.1
Thymus	0.3	Ovarian ca. OVCAR-4	0.0
Spleen	2.5	Ovarian ca. OVCAR-5	5.4
Lymph node	1.3	Ovarian ca. OVCAR-8	2.5
Colorectal	4.0	Ovarian ca. IGROV-1	2.7
Stomach	1.1	Ovarian ca.* (ascites) SK-OV-3	1.2
Small intestine	1.8	Uterus	3.8
Colon ca. SW480	0.0	Placenta	36.6
Colon ca.* SW620(SW480 met)	0.0	Prostate	2.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.2
Colon ca. HCT-116	0.0	Testis	1.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	1.4	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	1.5	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	11.0	Melanoma M14	0.0
Bladder	3.8	Melanoma LOX IMVI	0.0
Trachea	1.3	Melanoma* (met) SK-MEL-5	7.5
Kidney	7.4	Adipose	6.2

Table AKG. Panel 2D

Tissue Name	Rel. Exp (%) Ag815, Run 144791433	Tissue Name	Rel. Exp.(%) Ag815, Run 144791433
Normal Colon	5.8	Kidney Margin 8120608	7.0
CC Well to Mod Diff (ODO3866)	1.6	Kidney Cancer 8120613	100.0

CC Margin (ODO3866)	1.1	Kidney Margin 8120614	14.0
CC Gr.2 rectosigmoid (ODO3868)	0.6	Kidney Cancer 9010320	8.2
CC Margin (ODO3868)	0.8	Kidney Margin 9010321	24.7
CC Mod Diff (ODO3920)	0.3	Normal Uterus	6.6
CC Margin (ODO3920)	1.9	Uterus Cancer 064011	10.8
CC Gr.2 ascend colon (ODO3921)	1.8	Normal Thyroid	2.6
CC Margin (ODO3921)	0.7	Thyroid Cancer 064010	4.1
CC from Partial Hepatectomy (ODO4309) Mets	1.6	Thyroid Cancer A302152	2.8
Liver Margin (ODO4309)	1.3	Thyroid Margin A302153	2.9
Colon mets to lung (OD04451-01)	0.4	Normal Breast	3.5
Lung Margin (OD04451-02)	3.1	Breast Cancer (OD04566)	1.0
Normal Prostate 6546-1	2.5	Breast Cancer (OD04590-01)	2.3
Prostate Cancer (OD04410)	13.7	Breast Cancer Mets (OD04590-03)	3.7
Prostate Margin (OD04410)	10.4	Breast Cancer Metastasis (OD04655-05)	0.9
Prostate Cancer (OD04720-01)	5.8	Breast Cancer 064006	1.4
Prostate Margin (OD04720-02)	12.9	Breast Cancer 1024	1.4
Normal Lung 061010	3.2	Breast Cancer 9100266	1.3
Lung Met to Muscle (ODO4286)	0.6	Breast Margin 9100265	0.8
Muscle Margin (ODO4286)	1.0	Breast Cancer A209073	3.0
Lung Malignant Cancer (OD03126)	11.7	Breast Margin A209073	3.1
Lung Margin (OD03126)	5.0	Normal Liver	0.9
Lung Cancer (OD04404)	1.3	Liver Cancer 064003	1.0
Lung Margin (OD04404)	6.9	Liver Cancer 1025	0.8
Lung Cancer (OD04565)	0.7	Liver Cancer 1026	2.5
Lung Margin (OD04565)	3.2	Liver Cancer 6004-T	1.5
Lung Cancer (OD04237-01)	20.6	Liver Tissue 6004-N	0.3
Lung Margin (OD04237-02)	5.4	Liver Cancer 6005-T	1.4
Ocular Mel Met to Liver (ODO4310)	0.1	Liver Tissue 6005-N	0.5
Liver Margin (ODO4310)	1.1	Normal Bladder	3.8
Melanoma Mets to Lung (OD04321)	0.3	Bladder Cancer 1023	0.4
Lung Margin (OD04321)	12.2	Bladder Cancer A302173	2.2
Normal Kidney	81.2	Bladder Cancer (OD04718-01)	0.6
Kidney Ca, Nuclear grade 2 (OD04338)	22.5	Bladder Normal Adjacent (OD04718-03)	12.6
Kidney Margin (OD04338)	29.7	Normal Ovary	1.0
Kidney Ca Nuclear grade 1/2 (OD04339)	15.1	Ovarian Cancer 064008	7.7
Kidney Margin (OD04339)	46.3	Ovarian Cancer (OD04768-07)	0.4
Kidney Ca, Clear cell type (OD04340)	7.2	Ovary Margin (OD04768-08)	4.9

Kidney Margin (OD04340)	41.5	Normal Stomach	2.8
Kidney Ca, Nuclear grade 3 (OD04348)	5.5	Gastric Cancer 9060358	1.0
Kidney Margin (OD04348)	32.3	Stomach Margin 9060359	1.1
Kidney Cancer (OD04622-01)	4.6	Gastric Cancer 9060395	1.8
Kidney Margin (OD04622-03)	5.3	Stomach Margin 9060394	0.5
Kidney Cancer (OD04450-01)	12.6	Gastric Cancer 9060397	0.9
Kidney Margin (OD04450-03)	37.9	Stomach Margin 9060396	0.1
Kidney Cancer 8120607	0.9	Gastric Cancer 064005	1.9

Table AKH. Panel 3D

Tissue Name	Rel. Exp. %) Ag815, Run 164886712	Tissue Name	Rel. Exp.(%) Ag815, Run 164886712
Daoy- Medulloblastoma	1.9	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0
TE671- Medulloblastoma	0.3	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.4
PFSK-1- Primitive Neuroectodermal	0.4	Ramos- Stimulated with PMA/ionomycin 14h	0.5
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0
SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.3	Daudi- Burkitt's lymphoma	0.4
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	8.1
SK-N-SH- Neuroblastoma (metastasis)	39.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.3	RL- non-Hodgkin's B-cell lymphoma	0.5
Cerebellum	22.4	JM1- pre-B-cell lymphoma	0.0
Cerebellum	3.6	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	0.5	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	12.7	HUT 78- T-cell lymphoma	0.7
DMS-79- Small cell lung cancer	0.0	U937- Histiocytic lymphoma	0.7
NCI-H146- Small cell lung cancer	9.5	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	28.7	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	55.1	Caki-2- Clear cell renal carcinoma	2.0
NCI-H82- Small cell lung cancer	0.7	SW 839- Clear cell renal carcinoma	0.0

NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	71.2	Hs766T- Pancreatic carcinoma (LN metastasis)	53.6
NCI-H1299- Large cell lung cancer	0.2	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	28.5
NCI-H727- Lung carcinoid	7.2	SU86.86- Pancreatic carcinoma (liver metastasis)	4.6
NCI-UMC-11- Lung carcinoid	**100.0**	BxPC-3- Pancreatic adenocarcinoma	0.0
LX-1- Small cell lung cancer	0.0	HPAC- Pancreatic adenocarcinoma	10.2
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	15.6
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	6.1
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	51.1
NCI-H716- Colon cancer	70.2	T24- Bladder carcinma (transitional cell)	0.7
SW-48- Colon adenocarcinoma	0.0	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.3	HT-1197- Bladder carcinoma	0.5
LS 174T- Colon adenocarcinoma	0.4	UM-UC-3- Bladder carcinma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	0.0
NCI-SNU-5- Gastric carcinoma	0.1	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	1.1	SK-LMS-1- Leiomyosarcoma (vulva)	0.3
NCI-SNU-16- Gastric carcinoma	2.2	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0
NCI-SNU-1- Gastric carcinoma	0.6	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	1.7
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	1.7
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	2.1	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	27.2	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0

Table AKI. Panel 4D

Tissue Name	Rel. Exp.(% Ag815, Run 145703150	Rel. Exp.(%) Ag815, Run 145918553	Tissue Name	Rel. Exp.(%) Ag815, Run 145703150	Rel. Exp.(%) Ag815, Run 145918553
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	3.5	10.1
Secondary Th2 act	0.0	0.0	HUVEC IFN gamma	42.6	72.7
Secondary Tr1 act	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.6	3.2
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	2.1	5.3
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	10.3	22.4
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	15.0	31.9
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	3.3	16.2
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	58.6	68.3
Primary Tr1 act	0.0	0.0	Microsvasular Dermal EC TNFalpha + IL-1beta	5.9	13.9
Primary Th1 rest	0.0	0.4	Bronchial epithelium TNFalpha + IL1beta	0.1	0.0
Primary Th2 rest	0.2	0.7	Small airway epithelium none	0.2	0.0
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	0.1	0.0
CD45RA CD4 lymphocyte act	0.0	0.0	Coronery artery SMC rest	0.3	1.0
CD45RO CD4 lymphocyte act	0.0	0.0	Coronery artery SMC TNFalpha + IL-1beta	0.0	0.2
CD8 lymphocyte act	0.0	0.0	Astrocytes rest	0.0	0.0
Secondary CD8 lymphocyte rest	0.0	0.0	Astrocytes TNFalpha + IL-1beta	0.2	0.1
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.1	0.2
CD4 lymphocyte none	0.1	0.1	KU-812 (Basophil) PMA/ionomycin	0.0	0.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.1	0.0	CCD1106 (Keratinocytes) none	0.0	0.0
LAK cells rest	0.0	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0	0.0
LAK cells IL-2	0.0	0.0	Liver cirrhosis	2.9	8.9

LAK cells IL-2+IL-12	0.0	0.0	Lupus kidney	8.7	9.3
LAK cells IL-2+IFN gamma	0.0	0.2	NCI-H292 none	0.0	0.3
LAK cells IL-2+ IL-18	0.1	0.0	NCI-H292 IL-4	0.0	0.0
LAK cells PMA/ionomycin	0.0	0.0	NCI-H292 IL-9	0.0	0.0
NK Cells IL-2 rest	0.0	0.0	NCI-H292 IL-13	0.0	0.0
Two Way MLR 3 day	0.0	0.2	NCI-H292 IFN gamma	0.1	0.0
Two Way MLR 5 day	0.0	0.0	HPAEC none	15.6	24.7
Two Way MLR 7 day	0.0	0.0	HPAEC TNF alpha + IL-1 beta	2.8	5.0
PBMC rest	0.3	0.4	Lung fibroblast none	0.5	1.5
PBMC PWM	0.0	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0	0.2
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-4	0.3	0.7
Ramos (B cell) none	0.2	0.0	Lung fibroblast IL-9	0.0	1.0
Ramos (B cell) ionomycin	0.2	0.6	Lung fibroblast IL-13	0.5	0.3
B lymphocytes PWM	0.5	0.1	Lung fibroblast IFN gamma	0.1	0.7
B lymphocytes CD40L and IL-4	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
EOL-1 dbcAMP	0.2	2.0	Dermal fibroblast CCD1070 TNF alpha	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.5	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
Dendritic cells none	0.0	0.0	Dermal fibroblast IFN gamma	0.0	0.2
Dendritic cells LPS	0.0	0.0	Dermal fibroblast IL-4	0.1	0.0
Dendritic cells anti-CD40	0.0	0.0	IBD Colitis 2	0.3	0.5
Monocytes rest	0.0	0.0	IBD Crohn's	0.7	0.2
Monocytes LPS	0.0	0.0	Colon	1.0	5.3
Macrophages rest	0.0	0.0	Lung	6.9	14.7
Macrophages LPS	0.0	0.0	Thymus	**100.0**	**100.0**
HUVEC none	4.9	11.6	Kidney	1.1	3.0
HUVEC starved	16.0	30.6			

Table AKJ. Panel 5 Islet

Tissue Name	**Rel. Exp.(% Ag815, Run 254387842**	**Tissue Name**	**Rel. Exp.(%) Ag815, Run 254387842**

97457_Patient-02go_adipose	21.2	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	2.7	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	8.5	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	**100.0**	94712_Donor 2 AD - A_adipose	0.1
99167_Bayer Patient 1	0.5	94713_Donor 2 AD - B_adipose	0.4
97482_Patient-08ut_uterus	3.4	94714_Donor 2 AD - C_adipose	0.2
97483_Patient-08pl_placenta	39.5	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	0.2	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	3.1	94730_Donor 3 AM - A_adipose	0.2
97488_Patient-09pl_placenta	26.2	94731_Donor 3 AM - B_adipose	0.1
97492_Patient-10ut_uterus	9.2	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	89.5	94733_Donor 3 AD - A_adipose	0.3
97495_Patient-11go_adipose	8.5	94734_Donor 3 AD - B_adipose	0.2
97496_Patient-11sk_skeletal muscle	0.4	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	12.2	77138_Liver_HepG2untreated	0.0
97498_Patient-11pl_placenta	72.7	73556_Heart_Cardiac stromal cells (primary)	4.2
97500_Patient-12go_adipose	17.0	81735_Small Intestine	1.9
97501_Patient-12sk_skeletal muscle	1.0	72409_Kidney_Proximal Convoluted Tubule	0.9
97502_Patient-12ut_uterus	5.8	82685_Small intestine_Duodenum	0.6
97503_Patient-12pl_placenta	54.0	90650_Adrenal_Adrenocortical adenoma	1.1
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	2.1
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	6.7
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	1.4

AI_comprehensive panel_v1.0 Summary: Ag815 Highest expression of this gene is detected in control sample for ulcerative colitis (CT=27.6). This gene shows a widespread expression in this panel. Moderate to low levels of expression of this gene are detected in samples derived from normal and orthoarthitis/ rheumatoid arthritis bone, cartilage, synovium and synovial fluid samples, normal lung, COPD lung, emphysema, atopic asthma, asthma, allergy, Crohn's disease (normal matched control and diseased), ulcerative colitis (normal matched control and diseased), and psoriasis (normal matched control and diseased). Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders

including psoriasis, allergy, asthma, inflammatory bowel disease, rheumatoid arthritis and osteoarthritis.

The amp plot of another experiment (run 249247531) indicates that there were experimental difficulties with this run; therefore, no conclusions can be drawn from this data.

Panel 1 Summary: Ag127 Highest expression of this gene is detected in placenta (CT=25.4). High expression of this gene is also seen in testis and uterus. Therefore, therapeutic modulation of this gene may be useful in the treatment of reproductive disorders and fertility.

Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, melanoma, gastric, colon, lung, breast, ovarian, and brain cancers. Thus, therapeutic modulation of the expression or function of this gene or its protein product through the use of small molecule drug or antibodies may be effective in the treatment of pancreatic, gastric, colon, lung, breast, ovarian, and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adrenal gland, thyroid, pituitary gland, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 1.2 Summary: Ag815 Two experiments with same probe and primer are in good agreement. Highest expression of this gene is detected in placenta and cerebral cortex (CTs=24-25.6). In addition, expression of this gene is seen in brain, tissues with metabolic/endocrine functions such as pancreas, adrenal gland, thyroid, pituitary gland, heart, liver and the gastrointestinal tract, endothelial cells and in cancer cell lines derived from gastric, colon, lung, breast, ovarian, and brain cancers. This pattern correlates to

expression seen in panel 1. Please see panel 1 for further discussion on the utility of this gene.

Panel 1.3D Summary: Ag815 Highest expression of this gene is detected in cerebral cortex (CTs=27.4). In addition, expression of this gene is seen in brain, tissues with metabolic/endocrine functions such as adipose, pancreas, adrenal gland, thyroid, pituitary gland, heart, liver and the gastrointestinal tract, endothelial cells and in cancer cell lines derived from gastric, colon, lung, ovarian, and brain cancers. This pattern correlates to expression seen in panel 1. Please see panel 1 for further discussion on the utility of this gene.

Significant expression of this gene is also detected in fetal skeletal muscle. Interestingly, this gene is expressed at much higher levels in fetal (CT=29) when compared to adult skeletal muscle (CT=34). This observation suggests that expression of this gene can be used to distinguish fetal from adult skeletal muscle. In addition, the relative overexpression of this gene in fetal skeletal muscle suggests that the protein product may enhance muscular growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the GPCR encoded by this gene could be useful in treatment of muscle related diseases. More specifically, treatment of weak or dystrophic muscle with the protein encoded by this gene could restore muscle mass or function.

Panel 2D Summary: Ag815 Highest expression of this gene is detected in a kidney cancer (CT=28.3). Interestingly, expression of this gene is strongly associated with normal kidney samples as compared to kidney cancers. In addition, moderate to low levels of expression of this gene is also seen in colon, prostate, lung, breast, liver, bladder, ovarian, gastric and stomach cancers. Therefore, therapeutic modulation of this gene or its protein product through the use of antibodies and small molecule drug may be useful in the treatment of kidney, colon, prostate, lung, breast, liver, bladder, ovarian, gastric and stomach cancers.

Panel 3D Summary: Ag815 Highest expression of this gene is detected in a lung cancer cell line (CT=29.6). Moderate levels of expression of this gene is also seen in number of cell lines derived from lung, pancreatic, uterine, brain and colon cancers. Therefore, expression of this gene may be used as marker to detect the presence of these cancers. Furthermore, therapeutic modulation of this gene may be useful in the treatment of these cancers.

Panel 4D Summary: Ag815 Two experiments with same probe-primer sets are in good agreement. Highest expression of this gene is detected in thymus (CTs=27.7-28). Moderate levels of expression of this gene are also seen in endothelials cells including HUVEC, lung and dermal microvascular EC cells, and HPEAC cells. In addition, moderate to low levels of expression of this gene is also seen in liver cirrhosis, lupus kidney and normal colon, lung and kidney samples. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these endothelial cells and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, osteoarthritis and liver cirrhosis.

Panel 5 Islet Summary: Ag815 Highest expression of this gene is detected in placenta of a non-diabetic and obese patient (CT=28). Moderate levels of expression of this gene are mainly seen in placenta, uterus, adipose, kidney and small intestine of diabetic and non-diabetic patients. Please see panel 1 for further discussion on the utility of this gene.

AL. CG57209-02 and CG57209-03: EMR1 hormone receptor

Expression of gene CG57209-02 was assessed using the primer-probe set Ag6343, described in Table ALA. Results of the RTQ-PCR runs are shown in Tables ALB, ALC, ALD, ALE and ALF.

Table ALA. Probe Name Ag6343

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-caaataaataacatcttcagcgttct-3'	26	1003	362
Probe	TET-5'-cggtcgttttattttcacacactttgtcc-3'-TAMRA	29	1029	363
Reverse	5'-ctctcagttgtattcttcagagaaacta-3'	28	1058	364

Table ALB. AI comprehensive panel v1.0

Tissue Name	Rel. Exp.(%) Ag6343, Run 276596900	issue Name	Rel. Exp.(%) Ag6343, Run 276596900
110967 COPD-F	1.4	112427 Match Control Psoriasis-F	5.4
110980 COPD-F	2.2	112418 Psoriasis-M	2.5

110968 COPD-M	1.7	112723 Match Control Psoriasis-M	0.3
110977 COPD-M	6.7	112419 Psoriasis-M	3.4
110989 Emphysema-F	4.5	112424 Match Control Psoriasis-M	0.6
110992 Emphysema-F	2.2	112420 Psoriasis-M	14.2
110993 Emphysema-F	1.1	112425 Match Control Psoriasis-M	6.4
110994 Emphysema-F	2.1	104689 (MF) OA Bone-Backus	31.0
110995 Emphysema-F	8.4	104690 (MF) Adj "Normal" Bone-Backus	15.5
110996 Emphysema-F	0.5	104691 (MF) OA Synovium-Backus	3.9
110997 Asthma-M	3.8	104692 (BA) OA Cartilage-Backus	0.0
111001 Asthma-F	1.7	104694 (BA) OA Bone-Backus	9.2
111002 Asthma-F	1.9	104695 (BA) Adj "Normal" Bone-Backus	10.0
111003 Atopic Asthma-F	1.4	104696 (BA) OA Synovium-Backus	8.4
111004 Atopic Asthma-F	0.9	104700 (SS) OA Bone-Backus	**100.0**
111005 Atopic Asthma-F	0.4	104701 (SS) Adj "Normal" Bone-Backus	14.4
111006 Atopic Asthma-F	0.4	104702 (SS) OA Synovium-Backus	10.7
111417 Allergy-M	0.7	117093 OA Cartilage Rep7	5.5
112347 Allergy-M	0.0	112672 OA Bone5	23.7
112349 Normal Lung-F	0.0	112673 OA Synovium5	6.8
112357 Normal Lung-F	1.2	112674 OA Synovial Fluid cells5	12.2
112354 Normal Lung-M	0.9	117100 OA Cartilage Rep14	3.8
112374 Crohns-F	3.8	112756 OA Bone9	6.0
112389 Match Control Crohns-F	0.2	112757 OA Synovium9	0.7
112375 Crohns-F	6.1	112758 OA Synovial Fluid Cells9	4.5
112732 Match Control Crohns-F	17.4	117125 RA Cartilage Rep2	2.6
112725 Crohns-M	0.3	113492 Bone2 RA	42.6
112387 Match Control Crohns-M	1.6	113493 Synovium2 RA	14.9
112378 Crohns-M	0.0	113494 Syn Fluid Cells RA	26.8
112390 Match Control Crohns-M	1.9	113499 Cartilage4 RA	30.1
112726 Crohns-M	1.4	113500 Bone4 RA	29.9
112731 Match Control Crohns-M	1.9	113501 Synovium4 RA	18.2
112380 Ulcer Col-F	2.5	113502 Syn Fluid Cells4 RA	15.1
112734 Match Control Ulcer Col-F	43.2	113495 Cartilage3 RA	21.5
112384 Ulcer Col-F	10.1	113496 Bone3 RA	25.2
112737 Match Control Ulcer Col-F	1.6	113497 Synovium3 RA	11.7
112386 Ulcer Col-F	3.6	113498 Syn Fluid Cells3 RA	42.9
112738 Match Control Ulcer Col-F	8.8	117106 Normal Cartilage Rep20	0.3

112381 Ulcer Col-M	0.2	113663 Bone3 Normal	0.0
112735 Match Control Ulcer Col-M	0.8	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	1.2	113665 Syn Fluid Cells3 Normal	0.0
112394 Match Control Ulcer Col-M	0.7	117107 Normal Cartilage Rep22	0.8
112383 Ulcer Col-M	7.3	113667 Bone4 Normal	1.6
112736 Match Control Ulcer Col-M	0.0	113668 Synovium4 Normal	1.5
112423 Psoriasis-F	11.3	113669 Syn Fluid Cells4 Normal	1.4

Table ALC. CNS_neurodegeneration_v1.0

Tissue Name	**Rel. Exp.(%) Ag6343, Run 269225500**	**issue Name**	**Rel. Exp.(%) Ag6343, Run 269225500**
AD 1 Hippo	12.4	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	2.7	Control (Path) 4 Temporal Ctx	3.0
AD 3 Hippo	0.0	AD 1 Occipital Ctx	14.1
AD 4 Hippo	0.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	13.3	AD 3 Occipital Ctx	0.8
AD 6 Hippo	100.0	AD 4 Occipital Ctx	0.0
Control 2 Hippo	0.0	AD 5 Occipital Ctx	21.6
Control 4 Hippo	0.0	AD 6 Occipital Ctx	0.0
Control (Path) 3 Hippo	3.4	Control 1 Occipital Ctx	6.6
AD 1 Temporal Ctx	21.3	Control 2 Occipital Ctx	5.7
AD 2 Temporal Ctx	0.0	Control 3 Occipital Ctx	2.4
AD 3 Temporal Ctx	3.4	Control 4 Occipital Ctx	3.8
AD 4 Temporal Ctx	3.2	Control (Path) 1 Occipital Ctx	0.0
AD 5 Inf Temporal Ctx	0.0	Control (Path) 2 Occipital Ctx	0.0
AD 5 SupTemporal Ctx	14.1	Control (Path) 3 Occipital Ctx	6.5
AD 6 Inf Temporal Ctx	97.3	Control (Path) 4 Occipital Ctx	0.0
AD 6 Sup Temporal Ctx	50.0	Control 1 Parietal Ctx	10.7
Control 1 Temporal Ctx	2.8	Control 2 Parietal Ctx	6.9
Control 2 Temporal Ctx	1.2	Control 3 Parietal Ctx	10.0
Control 3 Temporal Ctx	12.9	Control (Path) 1 Parietal Ctx	0.0
Control 4 Temporal Ctx	0.0	Control (Path) 2 Parietal Ctx	3.5
Control (Path) 1 Temporal Ctx	0.0	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	2.9	Control (Path) 4 Parietal Ctx	0.0

Table ALD. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag6343, Run 259476287	issue Name	Rel. Exp.(%) Ag6343, Run 259476287
Adipose	12.2	Renal ca. TK-10	3.6
Melanoma* Hs688(A).T	0.0	Bladder	16.2
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	16.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	1.9
Prostate Pool	1.0	Colon ca. CaCo-2	0.0
Placenta	15.5	Colon cancer tissue	15.8
Uterus Pool	4.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	1.3	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	1.8
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	5.3
Ovary	4.1	Fetal Heart	3.8
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	3.3
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	1.1
Breast ca. T47D	0.0	Skeletal Muscle Pool	1.8
Breast ca. MDA-N	0.0	Spleen Pool	**100.0**
Breast Pool	3.0	Thymus Pool	22.1
Trachea	5.4	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	34.4	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	1.1	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	6.5
Lung ca. NCI-H23	0.0	Brain (fetal)	1.1
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	3.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	2.6
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	3.0

Liver	14.4	Brain (Thalamus) Pool	0.7
Fetal Liver	81.8	Brain (whole)	11.8
Liver ca. HepG2	0.0	Spinal Cord Pool	5.4
Kidney Pool	7.3	Adrenal Gland	12.1
Fetal Kidney	1.4	Pituitary gland Pool	2.4
Renal ca. 786-0	1.0	Salivary Gland	4.0
Renal ca. A498	1.0	Thyroid (female)	1.3
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.5	Pancreas Pool	4.9

Table ALE. Panel 4.1D

Tissue Name	Rel. Exp.(% Ag6343, Run 264776502	Tissue Name	Rel. Exp.(%) Ag6343, Run 264776502
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.1	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.3	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.2	HUVEC IL-11	0.0
Secondary Tr1 rest	0.1	Lung Microvascular EC none	0.0
Primary Th1 act	0.1	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.5	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.4	Microsvasular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.1	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.2	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.5	Coronery artery SMC rest	0.0
CD45RO CD4 lymphocyte act	1.4	Coronery artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.2	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.5	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.8	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.2	Liver cirrhosis	0.1

LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.1	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.3	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	0.2	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.9	HPAEC none	0.0
Two Way MLR 5 day	0.1	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	1.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	1.1	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.5	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.1	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	0.2	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	1.6	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.4	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.3	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	2.9
Monocytes rest	5.0	Neutrophils rest	17.0
Monocytes LPS	100.0	Colon	0.2
Macrophages rest	0.2	Lung	0.5
Macrophages LPS	1.4	Thymus	0.4
HUVEC none	0.0	Kidney	0.1
HUVEC starved	0.0		

<u>**Table ALF. Panel 5 Islet**</u>

Tissue Name	**Rel. Exp.(%) Ag6343, Run 259494665**	**Tissue Name**	**Rel. Exp.(%) Ag6343, Run 259494665**
97457_Patient-02go_adipose	45.1	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	55.5	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	13.8	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	61.1	94712_Donor 2 AD - A_adipose	0.0
99167_Bayer Patient 1	0.0	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0

97483_Patient-08pl_placenta	18.2	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	16.7	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.0	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	12.1	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	24.1	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	34.2	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	28.9	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	17.0	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	15.4	77138_Liver_HepG2untreated	0.0
97498_Patient-11pl_placenta	63.3	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	30.8	81735_Small Intestine	0.0
97501_Patient-12sk_skeletal muscle	15.3	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	21.6	82685_Small intestine_Duodenum	**100.0**
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	42.3
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.0
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

AI_comprehensive panel_v1.0 Summary: Ag6343 Highest expression of this gene is detected in orthoarthritis (OA) bone (CT=29.3). Low to moderate levels of expression of this gene are detected in samples derived from osteoarthritic (OA) bone and adjacent bone as well as OA cartilage, and OA synovial fluid samples. Moderate level expression is also detected in cartilage, bone, synovium and synovial fluid samples from rheumatoid arthritis patients. No significant expression of this gene is detected in normal samples of cartilage, synovium, bone or synovial fluid cells. Low to moderate level of expression is also seen in samples derived from COPD lung, emphysema, asthma, Crohn's disease (normal matched control and diseased), ulcerative colitis (normal matched control and diseased), and psoriasis (normal matched control and diseased). Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including psoriasis, allergy, asthma, inflammatory bowel disease, rheumatoid arthritis and osteoarthritis.

CNS_neurodegeneration_v1.0 Summary: Ag6343 Highest expression of this gene is detected in hippocampus sample derived from an Alzheimer's patient (CT=32.2). Moderate to low level of expression of this gene is alss seen in some of the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of this gene may be useful in the treatment of Alzheimer's disease.

General_screening_panel_v1.5 Summary: Ag6343 Highest expression of this gene is detected in spleen (CT=31.4). Moderate to low levels of expression of this gene is also seen in thymus, fetal lung and fetal liver. These tissues may contain monocytes or monocytic derived cell types. This gene codes for EMR1 hormone receptor precursor (human F4/80 homologue). EMR1 is a member of the family of hormone receptors with seven transmembrane segments. In addition, it has six egf-like modules at the N-terminus separated from the transmembrane segments by a serine/threonine-rich domain, a feature reminiscent of mucin-like, single-span, integral membrane glycoproteins with adhesive properties (Baud et al., 1995, Genomics 26(2):334-44, PMID: 7601460). EMR1 is shown to be abundantly expressed by cells of the myelomonocytic lineage (McKnight AJ, Gordon S., 1998, J Leukoc Biol 63(3):271-80, PMID: 9500513). A potential role for EMR3, a member of EMR family of proteins, has suggested in myeloid-myeloid interactions during immune and inflammatory responses. Therefore, therapeutic modulation of the EMR1 encoded by this gene through the use of antibodies directed against this molecule or a small molecule drug could inhibit monocyte activation or extravasation into inflamed tissue and may be important for the treatment of a number of inflammatory diseases including asthma and rheumatoid arthritis.

Among tissues with metabolic or endocrine function, this gene is expressed at low levels in adipose, adrenal gland, and liver. In addition, expression of this gene has been found to be dysregulated in CuraGen GeneCalling studies. It is upregulated in adipose tissue of mice who develop diabetes and obesity after being fed a high-fat diet. The EMR1 receptor encoded by this gene may be involved in a pathway leading to induction and release of TNF-alpha, IL-6 and resistin in adipose tissue. These molecules are known to be involved in the promotion of insulin resistance and are associated with obesity (Holst D, Grimaldi PA, 2002, Curr Opin Lipidol. 13(3):241-5, PMID: 12045392; Greenberg et al., 2002, Eur J Clin Invest. 32 Suppl 3:24-34, PMID: 12028372). Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of

endocrine/metabolically related diseases, such as obesity and diabetes, including Type 2 diabetes.

Interestingly, this gene is expressed at much higher levels in fetal (CTs=31.7-32.9) when compared to adult liver and lung (CTs=34-40). This observation suggests that expression of this gene can be used to distinguish fetal from adult tissues. In addition, the relative overexpression of this gene in fetal tissues suggests that the protein product may enhance liver and lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver and lung related diseases.

In addition, this gene is expressed at low levels in whole brain. Therefore, therapeutic modulation of this gene product may be useful in the treatment of neurological disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag6343 Highest expression of this gene is detected in LPS treated monocytes (CT=27.3). Expression of this gene is upregulated in activated monocytes as compared resting monocytes (CT=31.6). Therefore, expression of this gene may be used to distinguish between activated from resting monocytes and other samples used in this panel. The expression of this gene in LPS treated monocytes cells suggests that it plays a crucial role in linking innate immunity to adaptive immunity and also in initiating inflammatory reactions. Low to moderate levels of expression of this gene is also seen in neutrophils, eosinophils, PBMC, two way MLR, activated memory T cells, and CD4 lymphocytes. Therefore, modulation of the this gene or its product through the application of monoclonal antibodies or small molecule drug may reduce or prevent early stages of inflammation and reduce the severity of inflammatory diseases such as psoriasis, asthma, inflammatory bowel disease, rheumatoid arthritis, osteoarthritis and other lung inflammatory diseases. Please see panel 1.5 for further discussion on the utility of this gene.

Panel 5 Islet Summary: Ag6343 Low expression of this gene is restricted to sample derived from small intestine (CT=34.8). Therefore, expression of this gene may be used to distinguish this sample from other samples used in this panel. Please see panel 1.5 for further discussion on the utility of this gene.

AM. CG97715-01: TRANSMEMBRANE PROTEIN PT27

Expression of full-length physical clone CG97715-01 was assessed using the primer-probe set Ag3840, described in Table AMA. Results of the RTQ-PCR runs are shown in Tables AMB, AMC, AMD, AME and AMF.

Table AMA. Probe Name Ag3840

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-attcttagcagaatggggtgat-3'	22	693	365
Probe	TET-5'-cgctctcaactaactacaattgtattggca-3'-TAMRA	30	715	366
Reverse	5'-acaccataggggtcctctctag-3'	22	746	367

Table AMB. General screening panel v1.4

Tissue Name	Rel. Exp.(%) Ag3840, Run 217312795	issue Name	Rel. Exp.(%) Ag3840, Run 217312795
Adipose	18.9	Renal ca. TK-10	22.7
Melanoma* Hs688(A).T	53.2	Bladder	27.5
Melanoma* Hs688(B).T	56.6	Gastric ca. (liver met.) NCI-N87	59.9
Melanoma* M14	36.9	Gastric ca. KATO III	69.3
Melanoma* LOXIMVI	15.1	Colon ca. SW-948	22.2
Melanoma* SK-MEL-5	29.3	Colon ca. SW480	49.0
Squamous cell carcinoma SCC-4	19.3	Colon ca.* (SW480 met) SW620	30.1
Testis Pool	8.0	Colon ca. HT29	18.6
Prostate ca.* (bone met) PC-3	35.8	Colon ca. HCT-116	31.2
Prostate Pool	10.7	Colon ca. CaCo-2	24.1
Placenta	7.1	Colon cancer tissue	35.1
Uterus Pool	6.1	Colon ca. SW1116	12.4
Ovarian ca. OVCAR-3	26.6	Colon ca. Colo-205	15.8
Ovarian ca. SK-OV-3	52.9	Colon ca. SW-48	10.8
Ovarian ca. OVCAR-4	17.4	Colon Pool	14.4
Ovarian ca. OVCAR-5	50.7	Small Intestine Pool	8.8
Ovarian ca. IGROV-1	48.3	Stomach Pool	9.2
Ovarian ca. OVCAR-8	24.8	Bone Marrow Pool	7.6
Ovary	8.1	Fetal Heart	5.3
Breast ca. MCF-7	25.2	Heart Pool	7.2
Breast ca. MDA-MB-231	80.7	Lymph Node Pool	19.8
Breast ca. BT 549	73.2	Fetal Skeletal Muscle	3.6
Breast ca. T47D	100.0	Skeletal Muscle Pool	5.4
Breast ca. MDA-N	22.2	Spleen Pool	11.8
Breast Pool	14.2	Thymus Pool	12.7

Trachea	11.6	CNS cancer (glio/astro) U87-MG	51.4
Lung	3.0	CNS cancer (glio/astro) U-118-MG	81.2
Fetal Lung	16.5	CNS cancer (neuro;met) SK-N-AS	31.6
Lung ca. NCI-N417	12.8	CNS cancer (astro) SF-539	29.9
Lung ca. LX-1	21.3	CNS cancer (astro) SNB-75	61.6
Lung ca. NCI-H146	9.9	CNS cancer (glio) SNB-19	50.3
Lung ca. SHP-77	34.9	CNS cancer (glio) SF-295	61.6
Lung ca. A549	28.1	Brain (Amygdala) Pool	9.9
Lung ca. NCI-H526	12.9	Brain (cerebellum)	7.3
Lung ca. NCI-H23	30.4	Brain (fetal)	6.2
Lung ca. NCI-H460	17.9	Brain (Hippocampus) Pool	10.0
Lung ca. HOP-62	28.5	Cerebral Cortex Pool	9.8
Lung ca. NCI-H522	5.8	Brain (Substantia nigra) Pool	9.0
Liver	0.6	Brain (Thalamus) Pool	13.9
Fetal Liver	9.6	Brain (whole)	6.3
Liver ca. HepG2	7.8	Spinal Cord Pool	12.7
Kidney Pool	18.4	Adrenal Gland	14.1
Fetal Kidney	14.1	Pituitary gland Pool	3.7
Renal ca. 786-0	50.7	Salivary Gland	4.2
Renal ca. A498	13.8	Thyroid (female)	10.6
Renal ca. ACHN	12.5	Pancreatic ca. CAPAN2	48.0
Renal ca. UO-31	42.3	Pancreas Pool	33.4

Table AMC. Oncology cell line screening panel v3.1

Tissue Name	**Rel. Exp.(%) Ag3840, Run 223130227**	**Tissue Nme**	**Rel. Exp.(%) Ag3840, Run 223130227**
Daoy Medulloblastoma/Cerebellum	13.3	Ca Ski_Cervical epidermoid carcinoma (metastasis)	50.3
TE671 Medulloblastom/Cerebellum	11.7	ES-2_Ovarian clear cell carcinoma	20.3
D283 Med Medulloblastoma/Cerebellum	33.0	Ramos/6h stim_ Stimulated with PMA/ionomycin 6h	38.7
PFSK-1 Primitive Neuroectodermal/Cerebellum	36.1	Ramos/14h stim_ Stimulated with PMA/ionomycin 14h	20.4
XF-498_CNS	69.3	MEG-01_Chronic myelogenous leukemia (megokaryoblast)	75.8
SNB-78_CNS/glioma	38.7	Raji_Burkitt's lymphoma	13.8
SF-268_CNS/glioblastoma	35.4	Daudi_Burkitt's lymphoma	42.6
T98G_Glioblastoma	29.3	U266_B-cell plasmacytoma/myeloma	11.6
SK-N-SH_Neuroblastoma (metastasis)	37.4	CA46_Burkitt's lymphoma	14.6
SF-295_CNS/glioblastoma	41.5	RL_non-Hodgkin's B-cell lymphoma	9.5

Cerebellum	18.4	JM1_pre-B-cell lymphoma/leukemia	11.1
Cerebellum	10.0	Jurkat_T cell leukemia	24.0
NCI-H292_Mucoepidermoid lung ca.	99.3	TF-1_Erythroleukemia	**100.0**
DMS-114_Small cell lung cancer	7.5	HUT 78_T-cell lymphoma	28.3
DMS-79_Small cell lung cancer/neuroendocrine	13.0	U937_Histiocytic lymphoma	57.0
NCI-H146_Small cell lung cancer/neuroendocrine	35.4	KU-812_Myelogenous leukemia	66.9
NCI-H526_Small cell lung cancer/neuroendocrine	56.6	769-P_Clear cell renal ca.	39.2
NCI-N417_Small cell lung cancer/neuroendocrine	35.6	Caki-2_Clear cell renal ca.	28.1
NCI-H82_Small cell lung cancer/neuroendocrine	15.1	SW 839_Clear cell renal ca.	47.6
NCI-H157_Squamous cell lung cancer (metastasis)	37.6	G401_Wilms' tumor	17.0
NCI-H1155_Large cell lung cancer/neuroendocrine	49.3	Hs766T_Pancreatic ca. (LN metastasis)	50.3
NCI-H1299_Large cell lung cancer/neuroendocrine	26.2	CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)	33.4
NCI-H727_Lung carcinoid	61.6	SU86.86_Pancreatic carcinoma (liver metastasis)	52.5
NCI-UMC-11_Lung carcinoid	30.8	BxPC-3_Pancreatic adenocarcinoma	37.4
LX-1_Small cell lung cancer	37.9	HPAC_Pancreatic adenocarcinoma	74.7
Colo-205_Colon cancer	48.3	MIA PaCa-2_Pancreatic ca.	5.7
KM12_Colon cancer	71.7	CFPAC-1_Pancreatic ductal adenocarcinoma	92.7
KM20L2_Colon cancer	17.8	PANC-1_Pancreatic epithelioid ductal ca.	41.5
NCI-H716_Colon cancer	83.5	T24_Bladder ca. (transitional cell)	31.4
SW-48_Colon adenocarcinoma	31.2	5637_Bladder ca.	28.1
SW1116_Colon adenocarcinoma	12.7	HT-1197_Bladder ca.	51.4
LS 174T_Colon adenocarcinoma	20.7	UM-UC-3_Bladder ca. (transitional cell)	11.7
SW-948_Colon adenocarcinoma	24.8	A204_Rhabdomyosarcoma	28.9
SW-480_Colon adenocarcinoma	17.8	HT-1080_Fibrosarcoma	39.5
NCI-SNU-5_Gastric ca.	37.4	MG-63_Osteosarcoma (bone)	24.8
KATO III_Stomach	41.8	SK-LMS-1_Leiomyosarcoma (vulva)	71.7
NCI-SNU-16_Gastric ca.	18.2	SJRH30_Rhabdomyosarcoma (met to bone marrow)	32.5
NCI-SNU-1_Gastric ca.	75.8	A431_Epidermoid ca.	35.6
RF-1_Gastric adenocarcinoma	19.3	WM266-4_Melanoma	33.4
RF-48_Gastric adenocarcinoma	21.2	DU 145_Prostate	32.8
MKN-45_Gastric ca.	20.7	MDA-MB-468_Breast adenocarcinoma	27.9

NCI-N87_Gastric ca.	51.1	SSC-4_Tongue	18.9
OVCAR-5_Ovarian ca.	15.0	SSC-9_Tongue	37.4
RL95-2_Uterine carcinoma	20.4	SSC-15_Tongue	55.1
HelaS3_Cervical adenocarcinoma	33.9	CAL 27_Squamous cell ca. of tongue	21.8

Table AMD. Panel 4.1D

Tissue Name	Rel. Exp (%) Ag3840, Run 222546557	Tissue Name	Rel. Exp.(%) Ag3840, Run 222546557
Secondary Th1 act	41.2	HUVEC IL-1beta	64.6
Secondary Th2 act	41.5	HUVEC IFN gamma	42.6
Secondary Tr1 act	34.2	HUVEC TNF alpha + IFN gamma	40.6
Secondary Th1 rest	5.8	HUVEC TNF alpha + IL4	39.0
Secondary Th2 rest	9.2	HUVEC IL-11	20.7
Secondary Tr1 rest	5.7	Lung Microvascular EC none	77.4
Primary Th1 act	17.1	Lung Microvascular EC TNFalpha + IL-1beta	69.7
Primary Th2 act	33.7	Microvascular Dermal EC none	34.2
Primary Tr1 act	31.6	Microsvasular Dermal EC TNFalpha + IL-1beta	46.7
Primary Th1 rest	5.2	Bronchial epithelium TNFalpha + IL1beta	39.2
Primary Th2 rest	3.4	Small airway epithelium none	18.2
Primary Tr1 rest	12.8	Small airway epithelium TNFalpha + IL-1beta	69.7
CD45RA CD4 lymphocyte act	47.0	Coronery artery SMC rest	47.6
CD45RO CD4 lymphocyte act	36.6	Coronery artery SMC TNFalpha + IL-1beta	48.3
CD8 lymphocyte act	24.8	Astrocytes rest	25.5
Secondary CD8 lymphocyte rest	18.6	Astrocytes TNFalpha + IL-1beta	29.1
Secondary CD8 lymphocyte act	7.9	KU-812 (Basophil) rest	38.2
CD4 lymphocyte none	2.1	KU-812 (Basophil) PMA/ionomycin	57.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	11.8	CCD1106 (Keratinocytes) none	43.8
LAK cells rest	27.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	40.3
LAK cells IL-2	16.6	Liver cirrhosis	7.4
LAK cells IL-2+IL-12	15.3	NCI-H292 none	31.9
LAK cells IL-2+IFN gamma	10.7	NCI-H292 IL-4	40.6
LAK cells IL-2+ IL-18	19.8	NCI-H292 IL-9	50.3
LAK cells PMA/ionomycin	24.3	NCI-H292 IL-13	47.0

NK Cells IL-2 rest	20.9	NCI-H292 IFN gamma	39.8
Two Way MLR 3 day	29.1	HPAEC none	21.5
Two Way MLR 5 day	27.4	HPAEC TNF alpha + IL-1 beta	**100.0**
Two Way MLR 7 day	17.9	Lung fibroblast none	31.9
PBMC rest	2.3	Lung fibroblast TNF alpha + IL-1 beta	69.7
PBMC PWM	31.2	Lung fibroblast IL-4	33.4
PBMC PHA-L	30.4	Lung fibroblast IL-9	52.1
Ramos (B cell) none	26.6	Lung fibroblast IL-13	30.1
Ramos (B cell) ionomycin	40.1	Lung fibroblast IFN gamma	66.0
B lymphocytes PWM	20.3	Dermal fibroblast CCD1070 rest	54.7
B lymphocytes CD40L and IL-4	19.3	Dermal fibroblast CCD1070 TNF alpha	66.0
EOL-1 dbcAMP	31.9	Dermal fibroblast CCD1070 IL-1 beta	75.3
EOL-1 dbcAMP PMA/ionomycin	35.4	Dermal fibroblast IFN gamma	29.5
Dendritic cells none	39.5	Dermal fibroblast IL-4	39.0
Dendritic cells LPS	49.0	Dermal Fibroblasts rest	23.8
Dendritic cells anti-CD40	44.1	Neutrophils TNFa+LPS	3.1
Monocytes rest	18.9	Neutrophils rest	5.0
Monocytes LPS	90.1	Colon	5.4
Macrophages rest	35.8	Lung	27.2
Macrophages LPS	28.9	Thymus	10.4
HUVEC none	36.1	Kidney	16.5
HUVEC starved	42.0		

Table AME. Panel 5D

Tissue Name	**Rel. Exp.(%) Ag380, Run 169800718**	**Tissue Name**	**Rel. Exp.(%) Ag3840, Run 169800718**
97457_Patient-02go_adipose	27.5	94709_Donor 2 AM - A_adipose	**100.0**
97476_Patient-07sk_skeletal muscle	21.8	94710_Donor 2 AM - B_adipose	58.2
97477_Patient-07ut_uterus	24.5	94711_Donor 2 AM - C_adipose	52.9
97478_Patient-07pl_placenta	29.9	94712_Donor 2 AD - A_adipose	40.9
97481_Patient-08sk_skeletal muscle	39.2	94713_Donor 2 AD - B_adipose	48.6
97482_Patient-08ut_uterus	28.3	94714_Donor 2 AD - C_adipose	52.5
97483_Patient-08pl_placenta	32.1	94742_Donor 3 U - A_Mesenchymal Stem Cells	31.0
97486_Patient-09sk_skeletal muscle	8.3	94743_Donor 3 U - B_Mesenchymal Stem Cells	46.0

97487_Patient-09ut_uterus	44.8	94730_Donor 3 AM - A_adipose	94.6
97488_Patient-09pl_placenta	17.6	94731_Donor 3 AM - B_adipose	55.1
97492_Patient-10ut_uterus	47.6	94732_Donor 3 AM - C_adipose	55.1
97493_Patient-10pl_placenta	36.3	94733_Donor 3 AD - A_adipose	**100.0**
97495_Patient-11go_adipose	11.3	94734_Donor 3 AD - B_adipose	55.1
97496_Patient-11sk_skeletal muscle	7.4	94735_Donor 3 AD - C_adipose	66.9
97497_Patient-11ut_uterus	31.6	77138_Liver_HepG2untreated	45.7
97498_Patient-11pl_placenta	17.3	73556_Heart_Cardiac stromal cells (primary)	12.2
97500_Patient-12go_adipose	29.7	81735_Small Intestine	15.0
97501_Patient-12sk_skeletal muscle	11.6	72409_Kidney_Proximal Convoluted Tubule	22.1
97502_Patient-12ut_uterus	32.5	82685_Small intestine_Duodenum	14.1
97503_Patient-12pl_placenta	15.9	90650_Adrenal_Adrenocortical adenoma	20.2
94721_Donor 2 U - A_Mesenchymal Stem Cells	41.8	72410_Kidney_HRCE	64.2
94722_Donor 2 U - B_Mesenchymal Stem Cells	55.5	72411_Kidney_HRE	38.2
94723_Donor 2 U - C_Mesenchymal Stem Cells	37.9	73139_Uterus_Uterine smooth muscle cells	15.8

Table AMF. general oncology screening panel_v_2.4

Tissue Name	Rel. Exp.(%) Ag3840, Run 268036414	Tissue Nme	Rel. Exp.(%) Ag3840, Run 268036414
Colon cancer 1	35.6	Bladder cancer NAT 2	1.7
Colon cancer NAT 1	16.0	Bladder cancer NAT 3	2.6
Colon cancer 2	88.9	Bladder cancer NAT 4	4.5
Colon cancer NAT 2	16.5	Prostate adenocarcinoma 1	49.3
Colon cancer 3	77.4	Prostate adenocarcinoma 2	6.6
Colon cancer NAT 3	29.1	Prostate adenocarcinoma 3	24.3
Colon malignant cancer 4	**100.0**	Prostate adenocarcinoma 4	25.2
Colon normal adjacent tissue 4	8.5	Prostate cancer NAT 5	8.9
Lung cancer 1	45.7	Prostate adenocarcinoma 6	9.7
Lung NAT 1	4.7	Prostate adenocarcinoma 7	14.2
Lung cancer 2	75.8	Prostate adenocarcinoma 8	3.4
Lung NAT 2	8.4	Prostate adenocarcinoma 9	47.6
Squamous cell carcinoma 3	46.0	Prostate cancer NAT 10	5.3
Lung NAT 3	4.4	Kidney cancer 1	24.7
metastatic melanoma 1	14.2	KidneyNAT 1	7.6
Melanoma 2	4.4	Kidney cancer 2	62.9

Melanoma 3	4.8	Kidney NAT 2	23.5
metastatic melanoma 4	39.8	Kidney cancer 3	24.0
metastatic melanoma 5	50.3	Kidney NAT 3	4.1
Bladder cancer 1	7.5	Kidney cancer 4	18.4
Bladder cancer NAT 1	0.0	Kidney NAT 4	8.5
Bladder cancer 2	19.3		

General_screening_panel_v1.4 Summary: Ag3840 Highest expression of this gene is detected in a breast cancer T47D cell line (CT=25.3). High levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to high levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Interestingly, this gene is expressed at much higher levels in fetal (CT=28.7) when compared to adult liver (CT=32.7). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases.

Oncology_cell_line_screening_panel_v3.1 Summary: Ag3840 Highest expression of this gene is detected in a erythroleukemia TF-1 cell line (CT=26.6). This gene shows a widespread expression in all the cancer cell line and normal tissues in this panel. This pattern is in agreement with the expression profile in General_screening_panel_v1.4 and also suggests a role for the gene product in cell survival and proliferation. Please see panel 1.4 for further discussion on the utility of this gene.

Panel 4.1D Summary: Ag3840 Highest expression of this gene is detected in TNF alpha and IL-1 beta treated HPAEC cells (CT=27.8). This gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.4 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

Panel 5D Summary: Ag3840 Highest expression of this gene is detected in a midway differentiated and differentiated adipose tissue (CTs=29.4). This gene shows a widespread expression in this panel, which correlates to pattern seen in panel 1.4. Please see panel 1.4 for further discussion on the utility of this gene.

general oncology screening panel_v_2.4 Summary: Ag3840 Highest expression of this gene is detected in a malignant colon cancer sample (CT=26.6). Expression of this gene is seen in both normal and cancer samples derived from colon, lung, melanoma, bladder, prostate and kidney. Interestingly, expression of this gene is consistently higher in the cancer samples as compared to the corresponding normal adjacent tissues. Therefore, expression of this gene may be used as diagnostic marker to detect the presence of colon, lung, bladder, prostate and kidney cancers. Furthermore, therapeutic modulation of this

gene or its protein product may be useful in the treatment of colon, lung, melanoma, bladder, prostate and kidney cancers.

Example D: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions

identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn et al., Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. Genome Research. 10 (8) 1249-1265, 2000).

Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

NOV1a SNP data:

NOV1a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:1 and 2, respectively. The nucleotide sequence of the NOV1a variant differs as shown in Table 51A.

Table 51A data for NOV1a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381211	2786	T	G	829	Ile	Ser

NOV2b SNP data:

NOV2b has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:5 and 6, respectively. The nucleotide sequence of the NOV2b variant differs as shown in Table 51B.

Table 51B data for NOV2b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381047	516	T	C	148	Asn	Asn
13381110	1479	G	A	469	Gln	Gln
13381109	1542	C	T	490	Asp	Asp
13381108	1751	A	G	560	Asn	Ser
13381107	1821	C	T	583	Ile	Ile
13381106	3702	C	T	0		
13381105	3971	C	T	0		
13381104	4111	G	A	0		
13381103	4141	G	A	0		
13381102	4198	C	T	0		

NOV4c SNP data:

NOV4c has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:21 and 22, respectively. The nucleotide sequence of the NOV4c variant differs as shown in Table 51C.

Table 51C data for NOV4c						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380816	440	A	G	147	Ile	Val
13380815	511	A	G	170	Thr	Thr

NOV5b SNP data:

NOV5b has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:27 and 28, respectively. The nucleotide sequence of the NOV5b variant differs as shown in Table 51D.

Table 51D data for NOV5b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381095	372	C	A	97	Ser	Ser
13381096	465	C	T	128	Pro	Pro
13381097	1797	C	T	572	Cys	Cys
13381098	1845	T	C	588	Tyr	Tyr
13381062	2254	T	C	0		
13381063	2474	A	T	0		
13381100	2593	A	G	0		
13381101	2697	C	A	0		
13381064	3183	T	C	0		
13381065	3352	G	A	0		
13381066	3541	C	T	0		

NOV6b SNP data:

NOV6b has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:31 and 32, respectively. The nucleotide sequence of the NOV6b variant differs as shown in Table 51E.

Table 51E data for NOV6b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381083	168	T	C	6	Phe	Leu
13381202	181	T	C	10	Leu	Pro
13381084	359	C	A	69	Leu	Leu

13381085	539	A	G	129	Glu	Glu
13381086	545	G	A	131	Gln	Gln
13381087	566	C	T	138	Val	Val
13381088	658	A	T	169	Asn	Ile
13381092	786	T	C	212	Cys	Arg
13381093	908	T	C	252	Cys	Cys
13381094	933	T	C	261	Ser	Pro

NOV8b SNP data:

NOV8b has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:39 and 40, respectively. The nucleotide sequence of the NOV8b variant differs as shown in Table 51F.

Table 51F data for NOV8b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381053	770	G	A	257	Arg	Lys
13381052	965	C	T	322	Ser	Phe
13381051	1047	T	C	349	Gly	Gly

NOV10a SNP data:

NOV10a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:45 and 46, respectively. The nucleotide sequence of the NOV10a variant differs as shown in Table 51G.

Table 51G data for NOV10a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381212	700	C	T	193	Ser	Phe

13381213	1445	A	G	0		
13381214	1449	A	G	0		
13381215	1461	G	T	0		
13380817	1463	A	G	0		
13381217	1591	C	T	0		
13381218	1601	C	A	0		

NOV14b SNP data:

NOV14b has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:57 and 58, respectively. The nucleotide sequence of the NOV14b variant differs as shown in Table 51H.

Table 51H data for NOV14b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381055	323	G	C	102	Glu	Gln
13377369	324	A	G	102	Glu	Gly

NOV15a SNP data:

NOV15a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:59 and 60, respectively. The nucleotide sequence of the NOV15a variant differs as shown in Table 51I.

Table 51I data for NOV15a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381041	360	T	C	114	Cys	Arg

NOV17a SNP data:

NOV17a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:71 and 72, respectively. The nucleotide sequence of the NOV17a variant differs as shown in Table 51J.

Table 51J data for NOV17a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381195	38	C	A	0		
13381227	474	A	G	139	Thr	Thr

NOV20a SNP data:

NOV20a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:85 and 86, respectively. The nucleotide sequence of the NOV20a variant differs as shown in Table 51K.

Table 51K data for NOV20a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381060	1716	A	G	567	Pro	Pro

NOV21a SNP data:

NOV21a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:89 and 90, respectively. The nucleotide sequence of the NOV21a variant differs as shown in Table 51L.

Table 51L data for NOV21a

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381080	6069	T	G	2018	Asn	Lys
13381079	7885	G	A	2624	Asp	Asn
13381225	8295	C	T	2760	Phe	Phe
13381078	8365	A	G	2784	Asn	Asp

NOV24a SNP data:

NOV24a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:95 and 96, respectively. The nucleotide sequence of the NOV24a variant differs as shown in Table 51M.

Table 51M data for NOV24a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381045	439	T	C	78	Trp	Arg
13381262	736	A	G	177	Thr	Ala

NOV27b SNP data:

NOV27b has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:111 and 112, respectively. The nucleotide sequence of the NOV27b variant differs as shown in Table 51N

Table 51N data for NOV27b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381221	824	T	C	262	Phe	Ser

NOV28a SNP data:

NOV28a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:113 and 114, respectively. The nucleotide sequence of the NOV28a variant differs as shown in Table 51O

Table 51O data for NOV28a						
Variant	**Nucleotides**			**Amino Acids**		
	Position	Initial	Modified	Position	Initial	Modified
13381251	285	C	T	85	Cys	Cys
13381250	341	G	T	104	Gly	Val
13381249	501	C	T	157	Thr	Thr

NOV29a SNP data:

NOV29a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:117 and 118, respectively. The nucleotide sequence of the NOV29a variant differs as shown in Table 51P

Table 51P data for NOV29a						
Variant	**Nucleotides**			**Amino Acids**		
	Position	Initial	Modified	Position	Initial	Modified
13381050	406	G	T	95	Val	Val

NOV30a SNP data:

NOV30a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:119 and 120, respectively. The nucleotide sequence of the NOV30a variant differs as shown in Table 51Q

Table 51Q data for NOV30a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381049	1469	A	T	487	Gln	Leu
13381048	1857	A	G	616	Ile	Met

NOV32a SNP data:

NOV32a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:123 and 124, respectively. The nucleotide sequence of the NOV32a variant differs as shown in Table 51R

Table 51R data for NOV32a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381112	369	C	G	118	Ala	Ala

NOV32b SNP data:

NOV32b has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:125 and 126, respectively. The nucleotide sequence of the NOV32b variant differs as shown in Table 51S

Table 51S data for NOV32b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380823	113	A	G	23	Asn	Asp
13380824	1491	A	G	482	Tyr	Cys
13377028	1596	T	C	517	Val	Ala
13381208	1900	G	T	0		

13381207	2002	G	A	0		
13381206	2012	T	C	0		
13381205	2132	A	G	0		

NOV39b SNP data:

NOV39b has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:149 and 150, respectively. The nucleotide sequence of the NOV39b variant differs as shown in Table 51T

Table 51T data for NOV39b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381198	359	A	G	118	Ala	Ala
13381239	581	C	T	192	Leu	Leu
13381238	582	A	G	193	Asn	Asp
13381199	615	A	G	204	Lys	Glu
13381237	625	C	T	207	Ala	Val
13381236	631	T	C	209	Leu	Pro
13381235	705	G	A	234	Val	Met
13381234	714	A	G	237	Met	Val
13381232	777	T	C	258	Leu	Leu
13381231	821	G	A	272	Arg	Arg

NOV42a SNP data:

NOV42a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:155 and 156, respectively. The nucleotide sequence of the NOV42a variant differs as shown in Table 51U

Table 51U data for NOV42a

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381081	341	C	T	75	Arg	Cys
13381242	1661	G	A	0		
13381241	1678	C	T	0		

NOV43a SNP data:

NOV43a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:157 and 158, respectively. The nucleotide sequence of the NOV43a variant differs as shown in Table 51V

Table 51V data for NOV43a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381056	113	C	T	33	Thr	Ile
13381057	166	C	T	51	Leu	Phe
13381058	290	G	A	92	Gly	Glu
13381061	1485	T	C	490	Asp	Asp

NOV44a SNP data:

NOV44a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:159 and 160, respectively. The nucleotide sequence of the NOV44a variant differs as shown in Table 51W

Table 51W data for NOV44a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381043	319	A	G	75	Arg	Gly
13381075	351	A	G	85	Gly	Gly

13381074	603	T	C	169	Thr	Thr
13381073	862	C	T	256	Leu	Leu

NOV47d SNP data:

NOV47d has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:181 and 182, respectively. The nucleotide sequence of the NOV47d variant differs as shown in Table 51X

Table 51X data for NOV47d						
Variant	**Nucleotides**			**Amino Acids**		
	Position	Initial	Modified	Position	Initial	Modified
13381257	213	C	T	65	Pro	Pro
13375569	1316	C	T	433	Thr	Ile
13375568	1441	C	T	475	Arg	Cys
13375567	1545	G	A	509	Ala	Ala
13375566	1558	G	A	514	Asp	Asn
13375572	4235	A	G	1406	Tyr	Cys
13381256	4342	C	T	1442	Pro	Ser
13377613	4402	A	G	1462	Thr	Ala
13381255	4658	A	G	0		

NOV48c SNP data:

NOV48c has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:205 and 206, respectively. The nucleotide sequence of the NOV48c variant differs as shown in Table 51Y

Table 51Y data for NOV48c

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380257	118	A	G	40	Thr	Ala
13380253	842	T	C	281	Val	Ala
13380743	1435	C	A	479	Gln	Lys
13380741	1714	G	A	572	Val	Ile

NOV50a SNP data:

NOV50a has one SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:213 and 214, respectively. The nucleotide sequence of the NOV50a variant differs as shown in Table 51Z

Table 51Z data for NOV50a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381219	132	T	C	44	Pro	Pro
13375293	180	A	G	60	Thr	Thr
13381220	243	C	T	81	Ile	Ile
13374623	494	G	A	165	Gly	Asp
13375691	713	A	G	238	Asp	Gly

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or

library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims. The claims presented are representative of the inventions disclosed herein. Other, unclaimed inventions are also contemplated. Applicants reserve the right to pursue such inventions in later claims.

CLAIMS

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequenced selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107.

2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107.

3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107.

4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107.

5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.

6. A composition comprising the polypeptide of claim 1 and a carrier.

7. A kit comprising, in one or more containers, the composition of claim 6.

8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.

9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and

(c) determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.

10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and

b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

(a) introducing said polypeptide to said agent; and

(b) determining whether said agent binds to said polypeptide.

12. The method of claim 11 wherein the agent is a cellular receptor or a downstream effector.

13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

(a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;

(b) contacting the cell with a composition comprising a candidate substance; and

(c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.

14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

(a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;

(b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and

(c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator activity of or latency or predisposition to, a pathology associated with the polypeptide of claim 1.

15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to

a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.

19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107 or a biologically active fragment thereof.

20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 107.

21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.

22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 107.

23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 107.

24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of 2n-1, wherein n is an integer between 1 and 107.

25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group

consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 107, or a complement of said nucleotide sequence.

26. A vector comprising the nucleic acid molecule of claim 20.

27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.

28. A cell comprising the vector of claim 26.

29. An antibody that immunospecifically binds to the polypeptide of claim 1.

30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.

31. The antibody of claim 29, wherein the antibody is a humanized antibody.

32. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of said probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

33. The method of claim 32 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

34. The method of claim 33 wherein the cell or tissue type is cancerous.

35. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:

a) measuring the level of expression of the nucleic acid in a sample from the first mammalian subject; and

b) comparing the level of expression of said nucleic acid in the sample of step (a) to the level of expression of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of expression of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

36. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 107.

37. The method of claim 36 wherein the cell is a bacterial cell.

38. The method of claim 36 wherein the cell is an insect cell.

39. The method of claim 36 wherein the cell is a yeast cell.

40. The method of claim 36 wherein the cell is a mammalian cell.

41. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 107.

42. The method of claim 41 wherein the cell is a bacterial cell.

43. The method of claim 41 wherein the cell is an insect cell.

44. The method of claim 41 wherein the cell is a yeast cell.

45. The method of claim 41 wherein the cell is a mammalian cell.